U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 401

THE NMC/NCAR CDAS/ REANALYSIS PROJECT

R. Jenne, E. Kung, D. Salstein
E. Kalnay, M. Kanamitsu, R. Kistler
W. Collins, D. Deaven, L. Gandin, S. Saha
G. White, J. Woollen M. Chelliah, J. Janowiak
K. C. Mo, J. Wang A. Leetmaa, R. Reynolds

October 1993

This is an unreviewed manuscript, primarily intended for informal
exchange of information among NMC staff members

The NMC/NCAR CDAS/Reanalysis Project

E. Kalnay, M. Kanamitsu, R. Kistler,
W. Collins, D. Deaven, L. Gandin, S. Saha, G. White, J. Woollen
(NMC Development Division)

M. Chelliah, J. Janowiak, K. C. Mo, J. Wang
(NMC Climate Analysis Center)

A. Leetmaa, R. Reynolds
(NMC Coupled Model Project)

Roy Jenne
(NCAR)

E. Kung
(University of Missouri)

and

D. Salstein
(AER)

NMC Office Note 401

October 1993

Table of Contents

I. Introduction (Kalnay and Kanamitsu)

The NMC/NCAR Reanalysis Project began in 1991 as an outgrowth of the NMC Climate Data Assimilation System (CDAS) project. The motivation for the CDAS was the fact that the many changes introduced in the Global Data Assimilation System (GDAS) in order to improve the forecasts over the last decade, have also resulted in apparent "climate changes". These jumps in the perceived climate parameters obscure to some extent the signal of true short term climate changes or interannual climate variability.

The basic idea of the CDAS/Reanalysis Project is to use a frozen state-of-the-art analysis/forecast system and perform data assimilation using past data, from 1958 to the present (Reanalysis). Moreover, the same frozen analysis/forecast system will continue to perform data assimilation into the future (CDAS), so that climate researchers can assess whether current climate anomalies are significant compared to a long reanalysis without changes in the system.

The project development has been supported since 1990 by the NOAA Office of Global Programs, and an Advisory Panel (Julia Nogués-Paegle, Chair) has guided its design. The Project has the participation of scientists from the NMC Development Division, who have done most of the system development and testing, from NCAR, who has the responsibility for data collection, from the NMC Climate Analysis Center, who have participated in the design of the archive, the development of the automatic monitoring system and most of the diagnostics, from the NMC Coupled Model Project, who has provided the latest SST reanalyses and will perform the one-way coupled ocean reanalyses, from GFDL, who have contributed their own ocean reanalysis, from the NMC Automation Division, who have helped with longer data holds for CDAS/Reanalysis and with providing computer systems that allowed the performance of one month of reanalysis per day, and from the University of Missouri, who have collaborated with NCAR on the data quality control and acquired significant amounts of data from China not previously available (see Appendix A for a list of the personnel involved and their responsibilities). We have also had collaboration from NESDIS, who provided us with TOVS data, the UKMO who has promised to supply their SST and Ice reanalyses (GISST) for the earlier periods, the Japanese Meteorological Agency, who provided cloud- track winds and has offered special rawinsonde data not available on GTS, and ECMWF who has offered to fill some data gaps.

The early design of the Project was discussed in an NMC/NCAR Reanalysis Workshop that took place at NMC in April 1991 (Kalnay and Jenne, 1991) It had the participation of representatives of all the groups planning to perform reanalyses at that time (COLA, ECMWF and NASA/GLA), as well as of the major types of users (e.g., short and long term dynamics, transport of trace gases, climate change, predictability, angular momentum and length-of-day, coupled models, etc.). Since then, several other centers have planned or started a Reanalysis Project: The University of Maryland (Center for Ocean, Land and Atmosphere interactions, COLA) is performing an 18 month reanalysis for the 1982-82 El Niño. The European Centre for Medium-range Weather Forecasts (ECMWF) has a major project to reanalyze the period between 1979 and 1992. The NASA/GSFC Goddard Laboratory for Atmospheres (GLA) has started a 5-year Reanalysis project (Schubert et al, 1993). The US Navy has also indicated that it has plans for a 1985 to the present reanalysis. Such multiple projects for reanalysis offer a great opportunity for cooperation and intercomparison which should enhance each of the projects. In particular, NMC has benefitted from the COLA project through the transfer of the Gridded Analysis and Display System (GrADS), which has greatly enhanced the NMC graphical display system. NMC and ECMWF have agreed to provide missing data to each other for the 1985-1992 period, and NASA is considering the distribution of some of the NMC reanalysis products through its Mass Storage System.

Since the 1991 Workshop, the NMC/NCAR CDAS/Reanalysis System has undergone many major developments, so that the purpose of this paper is to provide an update of the system status and further plans before the operational implementation which will take place after the arrival of the Class 7 computer, in early 1994.

II. Basic characteristics of the NMC/NCAR Reanalysis System

Some of the characteristics of the first phase of the NMC/NCAR CDAS/Reanalysis are:

- T62 model (equivalent to a horizontal resolution of about 210 Km) with 28 vertical levels. The model is identical to the NMC model operational as of late 1993, except for the horizontal resolution, which is T126 for the operational model (Kanamitsu, 1990);
- SSI (3-D variational) analysis (Parrish and Derber, 1992, Derber et al, 1991);
- Complex QC (with corrections) of rawinsondes including time interpolation checks; OI-based complex QC of all other data;
- Optimal Averaging of a number of parameters over a number of areas (resulting in better averages and estimates of error of the average);
- OI SST reanalysis (latest analysis from R. Reynolds, starting from 1982);
- One-way coupled ocean model 4-D assimilation for 1982-onwards (from A. Leetmaa).
- The data input will be pre-processed, and all the analysis output fields will be post-processed with a "complex QC" monitoring system, in which the statistics of the data, time tendencies, etc, are compared to climatological statistics in order to detect errors. These statistics include tendency checks. (These automatic monitoring systems will be also implemented operationally, so that their development constitutes a major spin-off from this project for NMC).
- Output includes:

* Level-2 observational data in BUFR with QC information

* Analysis fields presented in "synoptic" form (every 6 hours) including basic variables and comprehensive diagnostic fields such as precipitation, surface fluxes and diabatic heating in sigma coordinates in GRIB format. A restart file will be included once a month to allow short reruns with extended diagnostics.

* "Time series" output ordered by day including standard pressure level fields, precipitation and other diagnostic fields. This format will be the most useful for most users.

* Possibly a special GRIB file for the North American region.

- Period covered by the Reanalysis: 1958 - 1993, with a continuation into the future through the CDAS. The reanalysis for the period 1985-1993 will be performed first, followed by 1979-1985, 1962-1979, 1958-1962. Reanalysis for the period 1946-1958, when the NH upper air network was established has been requested by atmospheric researchers (e.g., J. M. Wallace), and will probably be also carried out. The system has been designed to perform one month of reanalysis per day, once the systems have been thoroughly tested. The automatic QC monitoring systems for data preprocessing and output postprocessing are essential components needed to achieve this goal.

- 5-day forecasts every 5 days if the computer resources allow it.
- Expected completion of the first phase (35 years): 1996.

III. The Preparation of Data for Reanalysis (Jenne)

Surface and upper-air observations are being prepared for use in the planned reanalysis. The plan is to use the data available for the original operational NMC analyses (this data exists from Mar 1962-on), and to add other datasets to capture the older data from about 1948-on. Additional data inputs for 1962-on will provide much more data than was first available operationally.

Timing: The data for 1980-92 should be at NMC by at least Dec 1993, and the first version of the data for 1948-79 should be ready by mid-1994.

The component datasets are summarized in more detail below, and selected texts with more information are listed.

a) COADS Surface Marine Data

From about 1982-on there has been a large joint project involving NCAR, ERL (Boulder) and NCDC to prepare the best possible set of global ocean surface marine data. Canada and other countries have also helped. The Polar Science Center (University of Washington) has provided Arctic ice buoy data. The COADS dataset includes ships, fixed buoys, drifting buoys, pack ice buoys, near-surface data from ocean station reports (XBTs, etc.), and some other data. NCDC key-enters data from U.S. ship logs; data also comes in from many countries in delayed time. After a wait of 5 years COADS has about twice the number of ship data, compared to real-time data.

The data for 1854-1979 was released in 1983. A major 1980-92 update was completed in June 1993, after an intensive phase of 2 years. A copy of the recent data is being put into synoptic sort and sent to NMC (July 1993). Reynolds will then prepare the enhanced weekly 1° analyses of sea surface temperatures (1981-92) that are needed for reanalyses.

Additional data are now available for the older COADS period (1854-1979) and are being assembled. A revised COADS set for 1947-79 will be ready to send to NMC about April 1994 to use for reanalysis.

b) Global Rawinsonde Data

NCAR has tapes of the NMC GTS data with upper-air observations from Mar 1962-on. These are all original tapes from NMC, so no earlier conversions were made (only migration to new media). These data will be used for reanalysis. We plan to provide both the GTS data (which also has pibals and aircraft) and also raobs from national archives in various countries. A few days of GTS data have the wrong dates, which will be corrected. NCAR has raobs directly from some countries such as South Africa, Australia, Canada, and from the U.S. (NCDC). The USAF prepared a global collection of data (TD54) that is mostly for the period 1948-70, which will be included. GFDL is helping with processing and checking this set, which will all be ready for the first reanalysis. The University of Missouri (E. Kung) is helping with some of the checking between different sources of the same data.

NCAR, NCDC, Russia, Europe, and others (including WMO) have interests in improving the global archive of rawinsonde data. We anticipate various collaborations to improve the basic input sets and to accomplish merges. The results will be available for later reanalyses, not the first one. The collaboration should mainly be to prepare the component subsets of data correctly. Different methods will probably be used to accomplish the larger data merges of subsets.

c) Aircraft Data

Aircraft data are available from real-time GTS sources from NMC from Mar 1962-on. The data for 1962-72 was one of the types of data that NCAR received on 2000 tapes from NMC about 1973. Data from several other sources is being prepared, including USAF data (1947-59) and tropical aircraft data from Sadler. Both research and commercial aircraft for the tropical Atlantic GATE experiment (1974) will be available. We want the GATE analyses to be especially good, because there was also weather radar data that will help to check cumulus parameterization methods in models. Data from TWERLE constant-pressure balloons for the S. Hemisphere (Jul 1975 - Aug 1976) will be in the dataset. These balloons provide data like one level of a rawinsonde near 150 mb. We may be able to include aircraft data from New Zealand, where only half of the reports were ever transmitted over the GTS.

d) Surface Land Synoptic Data

Data from three main sources will be available by about May 1994 (or earlier) to provide fairly good coverage from 1949-on. These sources are the NMC decode of GTS data from Jul 1976-present; a dataset that R. Davis at NCDC prepared from the USAF GTS data for 1967-80 and a set of world synop data from the USAF (TD13) for earlier years. There are data sources to make further improvements, but they cannot be included in time for the first reanalysis runs.

e) Satellite Sounder Data (Kanamitsu and Jenne)

The basic radiances are available for the following:

SIRS IR sounders	Apr 1969 - Apr 1971
VTPR IR sounders	Nov 1972 - Feb 1979
TOVS sounders (IR, Vis, MSU, SSU)	Nov 1978 - present
HIRS data test	Aug 1975 - Mar 1976

The first three of these archives are at NCAR. The VTPR archives include basic data and the operational retrieved soundings. For TOVS, there is a separate archive of cloud-cleared radiances and associated retrievals (2.5° resolution). There are some large gaps in the 2.5° archive, but the archive of TOVS radiances is almost complete.

In the first phase of the Reanalysis we plan to use the original operational TOVS retrievals of NESDIS. The use of interactive retrievals which is planned to be implemented in operation sometime in late summer of 1993 will be tested in the reanalysis for short period and will be evaluated for possible use in the CDAS. Such a system would be used in the second phase of the CDAS/Reanalysis.

The data archived at NESDIS will be used for the period between 1979 and the present. There are several periods of missing data in the archive, and some of them extend over 40 days (in April-May 1986). Efforts are in progress to recover some of the missing data from NCDC archive through NCAR. We have talked with ECMWF about the possibility of recovering the missing data from raw radiances, although this requires a rather tedious cloud clearing process. It should be noted that the pilot experiments comparing reanalysis with and without the use of satellite data, to be discussed in Section IX, have provided useful information regarding the uncertainties of the analysis without satellite data. This is very important for the period before 1979 when no TOVS satellite soundings were available. With respect to the period before 1979, in principle we plan to assimilate VTPR data, at least for the Southern Hemisphere, although we have no recent experience with that data.

f) Satellite Cloud Drift Winds

These wind data from the original NMC tapes will be available for use. Some of the same (but sometimes more complete) data from national archives may also be included. In addition, we have received from the JMA their GMS cloud drift winds for the period 1978-1991.

g) Coverage of Data for Reanalysis

A text entitled "Data for Reanalysis: Inventories" (Jenne, about 115 pages, dated Nov 1992) has various maps and displays that give a feeling for the coverage of surface and upper air data that are already available. Most of this information covers the period from about 1948-on. The coverage of data is rather encouraging, even for the earlier years. We note, however, that rawinsonde observing networks for Antarctica and the west coast of South America did not start until July 1957.

h) Contribution from the University of Missouri (Kung/Jenne)

The data effort by the University of Missouri is focused in two areas: (1) quality assessment of the global upper-air data archive and sea surface temperature data in different periods from the late 1950s to the present, and (2) recovery of early upper-air data over poor coverage areas in the northern hemisphere.

The data quality examination is being made by checking the distribution and 4-dimensional consistency of available data, ratio and pattern of rejection by the hydrostatic check, and meteorological examination of large-scale patterns through EOF analysis and computation of kinematic fields.

The early daily upper-air data of 30 stations over China have been obtained from the Chinese NMC of the State Meteorological Administration for the period of 1958-1960. They will be edited and integrated into the current NCAR archive of upper-air data. The effect of this database integration will be carefully assessed.

Large quantities of upper-air data are also missing from the former Soviet Union, and an effort similar to the data retrieval from the Chinese mainland is in progress to recover them. Under the US-Russia bilateral exchange (effort led by Jenne and NCDC), the US has received (December 1990) 20 magnetic tapes with upper air data for 57 USSR stations for 1961-1978.

i) Use of hurricane data to generate synthetic observations (Lord, Kistler, Jenne)

We are currently trying to get the hurricane/typhoon data for the creation of tropical storm synthetic data through Steve Lord's procedure. The NHC can provide the data for the Atlantic and perhaps East Pacific hurricanes. Jenne is contacting the Joint Typhoon Center (Guam) who have sent such data to Charles Neumann, but there were problems with the data. Our aim is to get the hurrricane tracks at least for the latter period of reanalysis first (1985-93) and in the mean time consolidate and quality control the earlier data.

j) Other Information about Data for Reanalysis (Jenne)

Many reports have been prepared that give more information about the attributes of different datasets and the status of projects to prepare the data. Sea surface temperatures (SST) are needed for reanalysis; sea ice data is needed to prepare SST analyses. Tropical storm locations are needed also. Papers have been prepared that focus on different issues; a selection of these papers is given in Table 1.

Table 1. A Selection of Texts about Data for Reanalysis. Contact NCAR for further information.

Text	Date
Sea Surface Temperature data	1 Feb 1993
Sea Ice Data	2 April 1993
Data set of tropical storm locations	26 January 1993
TOVS Sounding data	25 January 1993
Missing data in the TOVS 2.5 Archive	1 March 1993
Analyses for the SH 1951-on	18 March 1993
NMC Upper Air Data, Inventories 1973-on	12 March 1993
Status of Reanalysis Data	1 April 1993
Ice Cap Buoys	4 June 1993

IV. Data Quality Control Preprocessor

The purpose of this Reanalysis module is to check the data (one year or more at a time) before the actual Reanalysis (one month/day) is performed. This should allow detection of major data problems with sufficient lead time (a few days before the execution of the Reanalysis), so that the human monitors can try to take corrective action. Hopefully this will minimize wasteful Reanalysis reruns needed because of the use of data with wrong dates, etc.

a) Satellite data (Kanamitsu)

A special satellite soundings data monitoring system independent from the complex quality control and the optimum interpolation quality control has been developed. It is aimed at quality controlling the data in the NESDIS archive tapes that can suffer from errors in dates and orbits not likely to occur in the daily operational products. The main technique followed for this purpose is the use of a climatology check and a tendency check similar to the ones performed in the analysis quality control to be discussed later. A group of satellite data in the grid box of 10 degree by 10 degree rather than a single satellite observation are quality controlled. The average in the box, the variance in the box and the absolute value of tendency of the box average are compared with the climatology to flag suspicious groups of satellite data.

b) CQC with temporal check (Collins)

The Complex Quality Control system is described in the next section. As indicated there, the QC preprocessor will include a version of the CQC without the use of the first guess of the model. It is possible that the baseline check (see next section) in the preprocessor will allow us to detect changes in the station locations before the Reanalysis is performed, which is also an important problem that interferes with the detection of climate change.

c) Climatological statistical QC test of data (Saha, Kistler, Kanamitsu)

The automatic monitoring system developed for the reanalysis output (Section VIII) has been so successful during the pilot tests, mostly because the climatological tests are based on grid point statistics, that we have decided to develop a test, using the same type of statistics, for the data preprocessor. The results of these tests will be expressed in units of standard deviation with respect to climatology, which will be added to the data in BUFR. This will have two effects: a) Allow human monitors to check for unusual data in unusual amounts, before the execution of the monthly reanalysis; b) Provide the OIQC with additional information for the Decision Making Algorithm as input to the reanalysis.

d) Fixed fields (Saha, Kanamitsu)

The best climatologies or latest reanalyses will be used as follows: 1) SST (Dick Reynolds after 1982 and the UKMO Global Ice and SST analysis (GISST) for earlier periods, David Parker); 2) Snow depth (NESDIS analyses and climatology, Brad Ballish); 3) sea-ice (the Joint Ice Center analyses when available, John Walsh and GISST analyses otherwise, Bob Grumbine); 4) albedo; 5) soil wetness (from SiB, Steve Schneider, Hua-lu Pan); (### Fiorino's suggestion: consider using Schemm et al, 1992 (NASA Tech memo 104571) soil moisture for 1979-88###) 6) Roughness length (from SiB); 7) resistance (from SiB). Monthly climatologies will be the back up of the analyzed fields when those are not available

All these fields will be compiled and put into a GRIB format. A simple QC checkout will be designed (comparison with monthly climatologies and standard deviations) to ensure that no major errors are present in the data or made inadvertently during their use.

V. Analysis module

a) Systems Configuration (Kistler)

i) Hardware

Cray YMP 8/64 MWord
File systems
/reanl1 - 10 GBytes
/reanl2 - 08 GBytes
Cray EL2
NMC NIC workstations (See distribution)

The CDAS/Reanalysis will be executed primarily on the machines of the NOAA Central Computer Facility in Suitland. When the project began, this site was primarily an IBM shop, with the now defunct Cyber 205 as a backend machine only accessible through the IBM clone HDS machines. When the Cray YMP8 was acquired in 1990, we began to move away from a centralized, hierarchical IBM type system, towards an open, distributed UNIX based system. One of the guiding principles the project has adopted was that all processing would be done in the UNIX environment. The other principle was that observations would be encoded in BUFR and gridded data in GRIB. In other words the project would take the lead in becoming independent of the antiquated IBM MVS operating system.

The first hardware configuration this principle modified from its original design involved the August 1992 connection of the STK robotic silo. Originally, the silo was planned to service MVS cartridge mounts and be managed by the existing Tape Management System. However, we proposed that the silo be dedicated to the 1500 tapes the CDAS/Reanalysis project had already purchased, and more importantly, that the Cray Data Migration Facility software be implemented with the Silo servicing the migration to cartridges. This move would break the existing Cray disk file gridlock, and move away from depending on MVS disk files for archiving purposes.

The second hardware configuration decision invoked by our Unix system principle involved the project's 18 Gigabyte disk space requirement. While large volumes of apparently inexpensive IBM 3380 disk packs were available, our accounting of total cost (avoiding wasted transfer time, avoiding needless pre-allocation via IBM JCL, avoiding future obsolescence, avoiding hours of people time maintaining backups, etc.) indicated that we would be better served with dedicated disk space on the Cray. When the Cray disk upgrade was finally performed in April 1993, we received the required space in two dedicated file systems: /reanl1 - 10GB and /reanl2 - 8GB.

The final configuration involved software. We anticipate writing about 10000 cartridges over the course of the reanalysis. This volume of tapes absolutely requires a modern tape management system. Again the only available system resided on the IBM system. Therefore we insisted on the purchase of the Cray Reel Librarian to service our cartridge management requirements. We expect this system to be brought online following the upgrade to UNICOS 7 during August, 1993.

Software

Unicos 7.x
NFS mount of Cray complex files
Bourne shell scripts
Mostly Fortran
a little C
very little X Windows
Data Migration Facility
Cray ReelLibrarian

Currently the NMC has been in the process of procuring a Class 7 computer. (IN early August 1993 it was announced that a Cray C-90 will be acquired. At the end of 1993 a 2 processor Cray C-90 will be installed, and the additional processors (total of 16) will be installed in March 1994. We have developed the CDAS/Reanalysis system to be run on the present Cray YMP, and the availability of the system is a requirement for the project. CDAS/Reanalysis will start when enough operational execution is transferred to the Class VII computer to allow its execution.

b) SSI (Kanamitsu)

The Spectral Statistical Interpolation, i.e., a 3-dimensional variational analysis scheme (Parrish and Derber, 1992, Derber et al., 1991) will be used as the analysis module. Of the several improvements to the original scheme that have been tested, two should be noted: a) the inclusion of the effect of surface friction, and b) better specification of forecast error and observational error statistics. The scheme is the same as used in the operational global data assimilation system as of 11 August, 1993. Since the pilot analysis demonstrated that the specification of satellite data observation error statistics influences the thermal structure of the analysis in the tropics and also the 5 day forecast skill to a large extent, several experiment will be performed to optimize the error statistics and improve the systematic bias of the analysis and the forecast before freezing the system for reanalysis production.

c) Model (Kanamitsu)

The T62/28 level NMC global spectral model will be used in the assimilation system. The vertical structure of the model is shown in Table 1. The model has 5 levels in the boundary layer and about 7 levels above 100 hPa. The lowest model level is about 5 hPa from the surface, and the top level is at about 3 hPa. This vertical structure was chosen so the boundary layer is reasonably well resolved and the stratospheric analysis at 10 hPa is not much affected by the top boundary conditions. The details of the model dynamics and physics are described in Development Division (1988), Kanamitsu (1989), and Kanamitsu et al (1990). The model includes parameterizations of all major physical processes, i.e., convection, large scale precipitations, shallow convection, gravity wave drag, radiation with diurnal cycle and interaction with clouds, boundary layer physics, a simple interactive surfacehydrology, and vertical and horizontal diffusion processes. A major difference with the model as described by Kanamitsu et al (1990) is the use of a simplified Arakawa-Schubert convective parameterization scheme developed by Hua-lu Pan at NMC based on Grell (1991). A large number of experiments were carried out during its development to ensure that the scheme behaves at least as well as the Kuo scheme used for many years. The experiments show that the Pan-Grell scheme results in much better prediction of precipitation over the continental US as measured by equitable threat scores. In addition, the precipitation patterns over the tropics are more realistic, with a smoother distribution, and less concentration over tropical orographic features.

d) Data Quality Control (Collins and Woollen)

NMC's operational quality control system has two components: i) The Complex Quality Control (CQC) system used for rawinsonde data, which because of the use of several independent checks is able to correct the majority of the errors. ii) The QC based on Optimal Interpolation (OIQC), which also performs multiple checks but does not attempt to correct data. The OIQC checks all the availble data is used immediately before the SSI analysis.

i) Complex Quality Control including temporal check (W. Collins)

The method of Complex Quality Control (CQC) (Gandin, 1988), is used to quality control the rawinsonde heights and temperatures for CDAS/Reanalysis. CQC first computes residuals from several checks (i.e., differences between an observation and the expected value from the check), and then uses an advanced Decision Making Algorithm (DMA) to accept, reject, or correct data (Collins and Gandin, 1990).

The checks used in the CQC code for rawinsonde heights and temperatures used for the CDAS/Reanalysis include: hydrostatic, baseline, incremental, horizontal, vertical and temporal. While significant level temperatures are also checked by the CQC code, their checking will not be described. It is understood in the following that mandatory level data only are under consideration.

The hydrostatic residual is the difference between the thickness of a layer (between two mandatory levels) computed from the reported heights and computed from the reported temperatures, using the hydrostatic equation. The baseline checks are based upon the difference between the station elevation and the elevation that is consistent with the reported surface pressure and the lowest two reported heights, using an assumed lapse rate of -6.5 degrees/km and the hydrostatic equation. Using the same information and assumptions, a mean-sea-level pressure may be obtained and compared with a forecast mean-sea-level pressure. In this way, both an increment and horizontal residual of mean-sea-level pressure is computed. The increment is the difference between a reported variable (height, temperature or surface pressure) and its short-term (6 hr) forecast value. The horizontal and vertical checks use increments in their computation. The horizontal residual is the difference between an increment (of height, temperature or surface pressure) and the increment interpolated horizontally from at most 4 neighbors, one from each quadrant. Finally, the vertical residual is the difference between an increment (of height or temperature) and the increment interpolated vertically from its nearest neighbor above and below. The interpolations are performed using optimal interpolation.

The CDAS/Reanalysis affords the possibility of performing another check--the temporal check, which cannot be done in the NWP system. The value of the heights and temperatures at observation time may be compared with those for 12 or 24 hours earlier and later. The temporal residual is the difference between the reported height or temperature and the value interpolated from one value before and one after, when they are available. Statistics show this check to be of comparable value to the incremental check. It is used along with any other checks that are available.

In addition, the CQC code diagnoses temperature observational errors. The hydrostatic residual does not react to these errors since the (erroneous) temperatures are used in computing the heights (also erroneous) at the observing site. These errors may be detected and rejected but not corrected.

As indicated before, there are two modules in the Reanalysis. The first is a data preprocessing module (described in this section IV) in which one or more years of data is checked without the use of a forecast first guess. In this stage, particularly, the temporal check is crucial. Without an increment with respect to the first guess, the incremental, horizontal and vertical residuals are not computed. The most powerful check of all, the hydrostatic check, is performed along with the baseline check and its preliminary diagnosis is confirmed or rejected by the temporal check.

In the second module, in which one month of Reanalysis is actually performed (see Section V), all checks are available, and the temporal check is normally used along with the other checks in making a decision. It is particularly useful here at locations that are isolated so that the horizontal check is not available.

The CQC for rawinsonde heights and temperatures performs quite well. The code has been running operationally at NMC for some years and has undergone steady improvement. At present, about 7% of the rawinsonde observations are found to have at least an error. Of the hydrostatically detectable errors in mandatory level heights and temperatures, 81% are corrected, and 60% of the errors detected by use of the baseline check are also corrected. The number of corrections for the early years of CDAS/Reanalysis may be anticipated to be somewhat smaller, depending upon data density, but there may be a higher percentage of observations that need to be corrected.

The CQC methodology is also used to quality control the mandatory and significant level rawinsonde winds. This is a new program not run operationally and initially it is designed only to identify erroneous winds. Later in its development, it will have limited capability to correct some wind errors, e.g. winds manually entered off by multiples of 100 degrees in direction or multiples of 100 knots in speed.

The wind CQC has the following checks for the wind components (u and v): incremental, horizontal optimal interpolation of increments, vertical optimal interpolation of increments, geostrophic and temporal. Particularly with the use of the temporal check, the wind CQC should have greater capability in detecting wind errors than other methods.

ii) OIQC (Woollen)

The Optimal Interpolation Quality Control (OIQC, Woolen, 1991) was developed as the final screening for observations to be used in the data assimilation. The goal of OIQC is to detect and withhold from the assimilation either of two types of undesirable observations. The first category contain gross errors caused by a variety of instrumental, human, or communication related mistakes which may occur during the process of making or transmitting observations. A number of automatic data preprocessing operations are performed, along with human monitoring, prior to the execution of the OIQC step. However, approximately 70-90% of the total observational database is not checked for any but the most obvious varieties of possible errors (i.e., those of impossible coordinates, completely garbled transmissions, etc.) before OIQC is executed.

The second category of problematic observations withheld by the OIQC contain large errors of representativeness. These observations are accurate, but their measurement represent spatial and temporal scales impossible to resolve properly in the analysis-forecast system. The assimilation of such data is likely to degrade the initial condition produced for the forecast model. The OIQC system uses the same statistical representativeness error model as the objective analysis system it precedes and, therefore, will detect observations which are unrepresentative to that system.

Three general principles are observed in the OIQC algorithm:

1) use of multivariate three-dimensional statistical interpolation for obtaining comparison values for each observation from nearby neighbors;

2) a complex of independent quality control components consisting of interpolation and other types of checks which when evaluated collectively suggest whether errors exist in an observation; and,

3) "non-hierarchical" decision making algorithm in which no final accept/reject decision is made for any datum until all checks which may affect that decision are completed.

The OIQC complex components are interpolation checks; an OI of the appropriate variable made to each datum, from a group of observations nearby, forms a comparison value. Univariate and geostrophic horizontal checks are performed for each datum checked, as well as a univariate vertical (profile) check for sounding data. Temperature data is converted to units of equivalent thickness difference from a background (usually a six hour forecast) for the checks, while wind data is judged in units of vector wind deviations.

A combination of individual check outcomes determines whether a datum is accepted by the system. The deviation of an observed value from an interpolation (o-a) is produced from each check. The deviations, normalized by expected values, are combined:

$$C^2 = \sum_{i=1}^{4} \frac{(o-a_i)^2 W_i}{(Va_i+Vo)} \bigg/ \sum_{i=1}^{4} W_i \quad . \qquad (1)$$

The weights (W) are formed from the normalized estimated reduction in the variance of each interpolation, calculated from the analysis and background variances:

$$W_i = 1 - Va_i/Vf \quad . \qquad (2)$$

The normalized combination is compared with a tolerance (T), empirically derived as a function of observation type. A mass observation passes the checks if

$$C^2 \leq T^2 \quad , \tag{3}$$

with wind data accepted when

$$C_u^2 + C_v^2 \leq T^2 \quad . \tag{4}$$

Checks of this nature are iterated four times to arrive at a sufficient convergence of decisions.

For the Reanalysis, we plan to add to the complex of checks performed by the OIQC, two more quantitative checks: One is produced by the "time tendency" check of the CQC (see previous section), and the second is the deviation with respect to climatology, measured in units of the local analysis standard deviation climatology (see the discussion in Section IV c). Both of these should be powerful additions to the QC, and are made possible by the use of the BUFR with QC system for storing data (see next subsection).

e) BUFR data formatting and QC information (Woollen, Kanamitsu)

The most crucial component of the Reanalysis is the level 2 data. The incoming data from various sources described in Section III will be first converted to the BUFR format, now the internationally accepted standard format for level 2 data. In addition, all the quality control information, i.e., radiation correction of rawinsondes, the flagging and correction of erroneous data by the Complex Quality Control of rawinsondes, the decisions of the OIQC and the final decision on the use of the data by the analysis (SSI) will be recorded into the BUFR data, following the BUFR formulation for QC data (see section VII.c). In this process, we plan to ensure that the original observation is not modified, and that corrections or changes are recorded as separate entries into the same observation. We also plan to encode the data increment with respect to the first guess into the same observation. As a result, and one of the most important outputs of the Reanalysis, the observational data base will have a common (BUFR) format, and QC information that should make it extremely useful in the evaluation of the analysis as well as in the future repetition of the reanalyses by NMC or by other centers.

It should be noted that the BUFR formatting with QC and the design of the BUFR archive is a new development designed for the Reanalysis Project. It is still undergoing the last stages of development.

f) Optimal Averaging (Gandin and Deaven)

Unlike the operational data assimilation systems, the NMC/NCAR Reanalysis system will include not only computation of grid-point values (and spherical harmonics), but also averaging over some prescribed areas, as well as in time. The incorporation of averaging procedures will result in increased possibilities of the global climate change detection because averaged values are less subjected to the small-scale everyday variability, that acts as a noise complicating the climate change signal detection.

A new method, known as the optimal averaging, will be used in the course of Reanalysis. This method assures minimum (in statistical sense) root-mean-square averaging errors and, what is particularly important for the Reanalysis purposes, it provides this rms error as a by-product.

Some work has been done at NMC in preparation to inclusion of the optimal averaging procedures into the Reanalysis system.

A rather simplified area optimal-averaging (OAv) model was designed allowing to perform the computations in a semi-analytical way and taking very little time (Gandin, 1993). Numerous computations with this model were performed to investigate the influence of various factors on the OAv weight pattern and particularly on its accuracy. The following main conclusions have been made: 1. Except for very small domains, the accuracy of OAv is substantially higher than that of the usually applied simple arithmetic averaging (AAv). This increase in accuracy is particularly high if deviations from the forecast first guess (increments), rather than those from climatology (anomalies) are averaged. In this case, however, knowledge of the accuracy of the first guess errors is necessary. 2. Although the OAv advantage over AAv is apparent even if data at only one point are used, this advantage is most pronounced when there are observations at several points. Under other equal conditions, the rms OAv error decreases with increasing observation density. However, if the density is already high, further increasing the data does not improve the OAv accuracy, but leads to a decrease in the accuracy of AAv. 3. The accuracy of OAv is very sensitive (although less than AAv) to violations of the symmetry of the observational pattern with respect to the averaging domain. 4. The OAv, like any averaging, possesses a kind of extremum properties. For example, with a given set of observation points, the rms averaging error is minimal for some area size and increases when the area is either larger or smaller. This increase is, however, much less pronounced for OAv than it is for AAv. 5. OAv of observations with correlated errors is less accurate and less computationally stable than that of data with non-correlated errors.

The first phase of the Optimal Averaging module has been coded and is currently being tested in the pilot version of the Reanalysis System. Optimal averages over prespecified areas are computed for temperature, specific humidity, u and v components of the wind, and wind speed at seven (1000, 850, 700, 500, 300, 200, 100 mb) pressures levels. The horizontal areas currently include nine 20 degree latitude bands from the south pole to the north and one hundred sixty two 20 by 20 latitude-longitude boxes covering the globe. The weights are computed by normalizing the optimal weights so that the sum of the weights over each area is equal to one. We plan to include also the geographical regions chosen by IPCC for climate change monitoring (IPCC 1990, p157, plus two regions covering South America: Tropical S.A., 10N-20S, 40W-80W; and Extratropical S.A., 20S-50S, 50W-70W).

Currently, the data input to the Optimal Averaging module is provided by the PREPDA file produced by the multivariate quality control program. Only data that have passed quality checks are used to obtain optimal averages. Optimal averages for each of the parameters over each area are archived in the form of ascii text files. The information can be scanned with display codes to obtain temporal averages or time series plots of each of the parameters.

Two additional computations will be included in the final version of the optimal averaging component. Optimal averages of the data increments (observation value minus first guess value) will be calculated by the same module that currently averages the actual observations. In addition, averages of the first guess over the same areas will be computed directly from the spectral coefficients of the assimilation model.

g) Periodic Forecasts from the Reanalysis (Kistler, Kalnay)

We have received a request from Prof. J. M. Wallace to perform daily global 5-day forecasts from the reanalysis. Although this would be a most useful addition to the data base, it would increase the computational effort by about 250%. Therefore, we will try to perform, instead, one 5-day forecast every 5 to 10 days, if feasible within the computational resources available. If such forecasts are performed, they would provide multiple benefits for: a) predictability studies; b) indirect estimates of the accuracy of the analysis; c) estimates of the impact of changes of the observing systems.

VI: The Climate Data Assimilation System (CDAS) (Kanamitsu, Kistler, Kalnay, Ropelewski)

This project originated with the idea of performing a "postanalysis" with a Climate Data Assimilation System (CDAS), which would remain frozen into the future (Ropelewski, pers. comm.). The CDAS analysis would be performed about a week after real time in order to a) capture the bulk of the delayed data, and b) be useful to the CAC in their generation of their monthly Climate Diagnostics Bulletins. When the Advisory Committee (in particular Profs. Mark Cane and Julia Nogués-Paegle) suggested that a long reanalysis would be much more useful than the CDAS alone, the Reanalysis component became the largest component of the project. It is clear that the combination of Reanalysis for the past, and a CDAS into the future, both using the same frozen system, is much more useful than either component alone. It should be noted that our plans include a second phase of the CDAS/Reanalysis sometime in 1997, after the first phase (CDAS/Reanalysis 1) is completed. In the second phase, the Reanalysis 2 would be started in 1979, and it would be performed with a 1997 state-of-the-art system, accompanied by a CDAS2 into the future. The reanalyses would then be repeated every five to ten years. The CDAS1, however, will be continued into the foreseeable future in order to maintain the longest homogeneous data assimilation product possible.

The CDAS will be performed with the same T62/28 level system of the reanalysis, and will be performed about a week after real time. This configuration calls for the following additions:

a) A copy of the level 2 data archive ("ten day archive") from the front end (NAS) computer system to the Cray (UNIX) system. This archive has a data hold of about 18 hours, which is enough to capture essentially all of the rawinsonde and surface data, except for ships and buoys. The plans of Automation Division call for the development of a new data base on the UNIX systems with a longer data hold but those will be available only about 18 to 24 months from now, and will then be used for CDAS.

b) An additional data dump for satellite data to avoid the present interorbital gaps produced by the operational dumps.

c) Use of the weekly SST analysis (see next section), which uses a whole week of both in situ and satellite data, and therefore should be centered at a time 3.5 days earlier than the analysis time. The one week delay of the CDAS with respect to real time should make this phase shift possible in practice.

d) Use of time interpolation checks for the complex quality control, also made possible by the 7-day delay in the execution of the CDAS.

VII. Reanalysis output (Kanamitsu, Janowiak, Saha, Woollen, Chelliah, Kalnay)

The design of the Reanalysis Output has been a major component of the Project. During the April 1991 Workshop it was clear that there were many different applications, and some of them, such as transport of greenhouse gases, have storage requirements that exceed what the project can handle (because of the need to save turbulent transports between any two layers). For this reason, it was decided that each unit of reanalysis output (one month) will include restart files, so that special purpose shorter reanalyses with extended output can be performed when needed.

The reanalysis archiving has been designed under the requirements that it should be both comprehensive, allowing, for example, the performance of detailed hydrological cycle budgets, and be easily accessible to the user interested in long series of data. It became clear that it was not possible to satisfy both requirements with a single archive format. For this reason, it was decided that the Reanalysis archive will have 3 major components:

a) Observations archive, with QC information, in BUFR format.

b) A comprehensive "synoptic" archive of analyses composed of several files containing sigma coordinate, pressure level and diagnostic fields. It is estimated that these fields, together with the observations will require about 10 cartridge tapes (0.2Gb each) per reanalysis month. It will also contain restart files. In addition, the data will also be stored in "time series" form in order to facilitate the creation of the following archive:

c) A reduced "time series" file allowing easy access by the user interested, for example, in getting only 500 hPa fields for many years. This archive will be saved by CAC as a subset of the archive b), and will be forced to fit in one cartridge per month.

In addition, and following a suggestion of Ralph Petersen, we are considering the creation of a special North American archive, possibly on a standard AWIPS grid, for the use of the NWS Forecast Offices and other interested US users. The 5-day forecasts requested by Prof. Wallace, if performed, would probably require an additional Lorenz' type of archive with a few pressure levels. Following a suggestion of Julia Nogués-Paegle, we are also working on the additional archiving of satellite data (e.g. OLR) with the same format as the other GRIB data.

The following table summarizes the type and volume of the two main archives:

REANALYSIS ARCHIVE VOLUME ESTIMATES

Table 1: Comprehensive "synoptic format" archives. See Appendix B for a detailed list of fields, units, etc., contained in each of the files listed below. Note: The actual number of tapes at NMC is between 2 and 3 times larger due to duplication of data in synoptic and time series format, and backup data.

FILE	ANAL	GUESS	TOTAL	MB/File	MB/day	GB/Mon	GB/35yr
Restart files (non-grib)							
Observation	4	0	4	2.0	8.0	0.24	102.2
Sigma	4	4	8	1.9	15.4	0.47	196.2
Surface	0	4	4	1.1	4.4	0.13	56.2
SST Anal	1	0	1	0.3	0.3	0.01	3.3
Snow/Ice	1	0	1	0.3	0.3	0.01	3.3
Grib files (Grid point)							
Pressure	4	4	8	1.4	11.4	0.35	145.1
Sigma	4	4	8	2.9	23.5	0.72	300.7
Diag.f06	0	4	4	5.8	23.2	0.71	296.4
Diag.f00	4	0	4	0.1	0.4	0.01	5.1
Fluxes	0	4	4	0.5	2.2	0.07	27.6
Non-grib file							
Zonal	4	4	8	0.0	0.2	0.01	3.1
Total (GB)						2.00	840
STK cartridges (.2GB each)						10	4200

REDUCED "TIME SERIES" ARCHIVE (Janowiak, Saha, Chelliah)

The CAC plans to save basic upper air parameters on standard pressure levels (Table 2), selected surface flux fields (Table 3), and diabatic heating and radiation terms for each analysis cycle for the entire Reanalysis period. Most of the data will be saved in GRIB format. The pressure level data will be saved on a 2.5 degree lat/lon grid, while the surface flux fields and radiation/diabatic heating data will be saved on a T62 Gaussian grid (192 x 94). The radiation/diabatic heating data will be composed of appropriate terms that will be combined to yield a single net radiative term and a single net heating term for each grid location. These data will be stored for each sigma level.

The data storage order will be markedly different from the manner in which model data have traditionally been stored at NMC, and more similar to the storage used for the NMC DERF experiments. Since the research community (particularly the climate research community), generally use individual parameters at a single atmospheric level but for many time periods (rather that all parameters for a single time), much of the data will be stored in chronological, not synoptic order. That is, individual fields for a single atmospheric level will be available for "all time" from a single data structure. The basic pressure level data and surface flux data will be stored in this order. The radiation/diabatic heating fields will stored in the traditional (synoptic) manner.

An estimate of the number of the number of storage units (STK cartridges) required for the 1957-92 period (35 years) follows:

Pressure level data set : 332 for 35 years/66GB
Surface flux data set : 120 for 35 years/24GB
Diabatic heating data set: 260 for 35 years/52GB

ESTIMATED TOTAL: 712 for 35 years/142GB

Table 2. Data fields on standard pressure levels to be saved on a 2.5 degree lat/lon grid (144 x 73).

U-component of wind
V-component of wind
Geopotential
Temperature
Vertical Velocity
Specific Humidity
Precipitable Water

16 levels: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10 mb

Table 3. Surface flux data to be saved on a T62 Gaussian grid (192 x 94):

Surface Pressure	Convective precipitation
Surface Temperature	Total precipitation
Soil Wetness	Surface heat flux
Snow Depth	Surface latent heat flux
U-component @ 10 meters	Ground heat flux (land only)
V- " " " "	Surface U & V stress
Sigma level 1 temperature	Surface longwave flux
" " 1 specific humidity	Surface shortwave flux
Cloud fraction (hi, mid, low)	Outgoing flux (short, long)

BUFR Observational file (J. Woollen)

The observation dataset undergoes multiple processing stages that can influence the quality of subsequent analysis and forecast products. The NMC Reanalysis system has several quality control steps (see sections IV and V, on the Preprocessor and Analysis modules), in which the data are compared with other baseline estimates (i.e. climatology, short forecasts, or other observations close in time and space). In addition, the Complex QC performs actual corrections of the observations, which are also modified through a "Radiation correction" that takes into account the effect of direct sunshine on thermistor measurements. Human intervention in the quality control process is also carried out in many operational centers, resulting in human judgements which override automatic review, or in modifications or additions to the observation data itself. In the reanalysis system this intervention will be limited to the monitoring of the results of the Data Preprocessor, and consequent changes required before the actual reanalysis is carried out.

For purposes of operational monitoring and performance review, and for research based on the Reanalysis, it would be helpful to be able to trace all actions and reactions (denoted "events") taken by any QC procedures on any particular observation, or set of observations. This could be a daunting task unless some summary of available information is maintained as a standard output from the system. Instead of creating a separate quality control "events file" for the NMC/NCAR reanalysis system, we decided to design a BUFR format for the observational dataset which allows recording the QC information, accumulating it from one step to the next, and finally, transferring this information, imbedded within the observation dataset, as part of the Level 2 output.

The method adopted for storing QC events in the reanalysis datasets can be briefly described, given a basic understanding of the structure of the observational dataset. It is useful, in this context, to stratify units of storage within the observation dataset into reports, observations, and measurements. We define a report as one or more atmospheric measurements, taken at one or more four dimensional locations, by a single observing platform or device (e.g., a rawinsonde report). An observation is a group of measurements from the same 4D location (e.g., height, temperature, humidity and wind components measured at a mandatory pressure level). At present, each report in the operational reanalysis dataset has fixed latitude, longitude, and time stamps, with one or more observations varying by pressure; thus several vertical sounding observations are contained within one report (the issue of spatial and temporal variation within balloon soundings or dropsondes is not addressed), whereas observations from horizontally moving platforms such as ships or aircraft are contained in multiple reports.

Each measurement in each observation of any report may undergo several quality control actions or reactions (events). If we recorded each event after each each individual measurement, the amount of information needed to cross reference the event to the measurement would be minimal. However, we have chosen to archive the events at observational boundaries by means of "push down" stacking, because it gives a useful chronology of event production at a rather small cost. Each event includes four coded items. The first is an indication of which QC program or procedure caused the event. Next, a standardized QC decision code indicates whether a measurement is judged to be usable or not, and attaches some level of confidence and/or precedence regarding the judgement. Third, a set of information specific to a particular QC process, which would be relevant to understanding the judgement made by that process (e.g., difference between the measurement and the first guess). This information will vary depending on the QC process that creates it, and its content is defined by the developer or maintainer of the QC process, based on the components and possible outcome of that process. The final item is reserved for storing the previous value of a measurement, for those processes which may modify the measurement itself.

The logistics involved in automatically recording and reviewing QC events are quite challenging. The flexibility of the BUFR format makes the methodology described above feasible. Manipulation of the BUFR reanalysis datasets has been simplified by the development of databasing BUFR interface program library routines, which separate application programmers from many of the technical aspects of the BUFR format, allowing read/write access to information by means of mnemonic reference. The machine independent FORTRAN programs and documentation will be available to investigators.

VIII. Automatic monitoring system

a) Automatic QC monitoring system for the Reanalysis output (S. Saha and M. Chelliah)

Once operational Reanalysis at NMC begins, it is expected that one month of analyses (analyzed and archived every 6 hours) will be produced in one calendar day. Thus, a time series of approximately 124 (4 X 31) "cycles" will be available at the end of each day. Since the volume of output data is huge, it is impossible for a human to check that the reanalysis products are correct, and detect major errors, drift, etc. In order to fulfill this important requirement we have developed an automatic monitoring system. We check the 6-hourly times series of geopotential height, zonal wind, meridional wind and temperature at 12 standard pressure levels (i.e. 12 X 4 = 48 variables at 124 time levels) at the end of each analysis month. These time series are obtained by converting the T62 model spectral coefficients to a 2.5° x 2.5° regular latitude/longitude grid. Time series of surface flux variables every 6 hours are also included in our monitoring system.

Climatology

To effectively monitor the pressure time-series, we use a climatology based on NMC's daily global data assimilation system (GDAS) covering a 7-year period from 1 July 1986 to 30 June 1993. From these daily values we computed monthly averages, standard deviations from the monthly means as well as standard deviations between successive analysis (at 00Z) for geopotential, zonal wind, meridional wind and temperature at each grid point at 12 mandatory pressure surfaces; definitions are given below.

For monitoring the surface flux quantities, a climatology based on 1 year (1 Feb 1992 - 31 Jan 1993) of daily surface flux files from a T62 model-based GDAS system at NMC has also been made.

The climatological parameters computed are as follows :

1. Monthly climatological mean for any variable over the 7-year daily data set for each grid point is :

$$\overline{C_{i,j,m}} = \frac{\sum_{t=1}^{t=7x\,n} C_{i,j,t}}{7x\,n}$$

where i,j are grid point indices, t is time (unit of days), n is the number of days in the month, and m is the index for calendar month.

2. Climatological daily standard deviation as a function of month for any variable over the 7-year daily data set for each grid point is :

3. Monthly climatological standard deviation of the tendency between daily observed values over the 7 year data set for each grid point is :

$$sdC_{i,j,m}=\sqrt{\frac{\sum_{t-1}^{t-7x\,n}(C_{i,j,t}-\overline{C_{i,j,m}})^2}{7x\,n}}$$

$$sd\Delta C_{i,j,m}=\sqrt{\frac{\sum_{t-1}^{t-7x\,n}(C_{i,j,t}-C_{i,j,t-1})^2}{7x\,n}}$$

Automatic Checks:

i) The "OUTLIER" check consists of computing $Q^o_{i,j,t}$ which is defined by

$$Q^o_{i,j,t}=\frac{(X_{i,j,t}-\overline{C_{i,j,m}})}{sdC_{i,j,m}}$$

where X is the data to be checked at grid point i,j at time level t. For each time-level we sum up the points for which $|Q^o_{i,j,t}|\geq 2.0$ and divide by the total number of grid points. If this ratio is greater than the highest value obtained from our 7-year climatology we suspect something could be wrong and proceed with further checks and diagnostics.

Furthermore we also check for extreme outliers by targeting grid points for which $|Q^o_{i,j,t}|\geq 10.0$ This check has proved to be very important in identifying bad input data, such as radiosonde data or satellite wind data.

ii) The "TENDENCY" check consists of computing the data tendency between successive time levels as:

$$\Delta X_{i,j,t}=X_{i,j,t}-X_{i,j,t-1}$$

and then computing $Q^t_{i,j,t}$ which is defined by:

$$Q^t_{i,j,t}=\frac{\Delta X_{i,j,t}}{sd\Delta C_{i,j,m}}$$

For each time level we sum up all grid points for which

$|Q_{i,j,t}^{t}| \geq 2.0$ and divide by the total number of grid points. If this ratio is greater than the highest value obtained from our 7-year climatology we suspect something could be wrong and proceed with further checks and diagnostics.

Furthermore we also check for extreme outliers by targeting grid points for which

$|Q_{i,j,t}^{t}| \geq 10.0$ This check has also proved to be very important in identifying bad input data.

b) An example of the application of the automatic monitoring:

The automatic monitoring system was further strengthened by the development of a highly sophisticated and fast graphical display capability of suspect fields. This entire suite of programs has been tested on the pilot experimental analyses for July and August 1985, and proved to be robust in detecting several errors, which have now been corrected in the reanalysis system.

An example of one such error is illustrated in Figure VIII(a-h). Figure a shows the full 500 mb temperature field on 23 July 1985 at 18Z. The monitoring system found a problem in the Caribbean (shown in the hatched box). Figures b and c show the data for this date and the climatology for this month respectively, for the hatched box region. At four grid points (marked by X's) the difference between the data and the climatology is clearly unacceptable (-8°C) as pointed out in Fig. d. The standard deviation allowed by the climatology (0.7°C) is shown in Fig. e. As can be seen from Fig. f, the ratio of the difference to the climatological standard deviation exceeds 10 at the X's. Figures g (non-satellite) and 1h (satellite) show the available data for the analysis at this time. Figure 1g shows the temperatures reported from aircraft along a track from Louisiana, east of the Yucatan onto Mexico are all around -18°C, which obviously are the culprits. While supporting each other in a buddy-check, they are about 11°C colder than the satellite-retrieved temperatures, as shown in Fig. h. It turned out that although AIREPS are not allowed to confirm each other, these were reconnaissance reports, whose data are encoded with an increasing number for each subsequent report. As a result the OIQC system assumed they had different airplane ID's. This subtle potential error has now been eliminated from the OIQC.

The above case was a subtle but easy problem that could be solved. Some of the more difficult problems are discussed below:

1. Unprecedented events, such as tropical storms, may set off alarm bells due to the fact that none of them occurred at the same grid point during the 7-years that make up our climatology. This problem may be alleviated by using a historical record of hurricane occurrences, which is being created at NMC from NCAR data.

2. Problems occur when slow and incremental errors associated with model drift enters the analysis repeatedly through the first guess. In particular, the pilot experimental reanalysis for July and August 1985 showed that when the reanalysis is restarted with a very different first guess (in this case the 1985 operational analysis) resulted in about two weeks of spinup of the system in the tropical regions.

Figures for Section VIII: Automatic Monitoring Program

Fig. VIIIa. Analyzed Temperature T at 500 mb on July 23, 1985 at 18z from Pilot 2 experimental Reanalysis. The problem region, as detected by automatic monitoring, is shown hatched.
Fig. VIIIb. Temperature over the hatched regions in Fig. VIIIa shown in expanded scale. The x's denote points which failed the monitoring test.
Fig. VIIIc. Climatology of temperature over the hatched region in Fig. VIIIa.
Fig. VIIId. Data - climatology
Fig. VIIIe. Standard deviation of temperature.
Fig. VIIIf. Ratio of (Data - climatology) to the standard deviation of temperature.
Fig. VIIIg. Non-satellite data that went into the analysis.
Fig. VIIIh. Raw satellite (TOVS) data that went into the analysis.

c) Monitoring of the data during the analysis (Kistler, Woollen)

A number of montoring plots were designed to allow a human monitor to check the enormous amounts of one month level 2 reanalysis data data easily in order to detect serious problems. In addition to the climatological check of the data described in the Data Preprocessor Section IV, and the normal operation of the OIQC and the CQC, after the execution of the Reanalysis for the month, the following statistics are plotted for human monitoring:

Data density plots for all observations (fig. VIII 1), notice the presence of Automatic Ship Aerological Program (ASAP) data at e.g., 30 N. Raw (solid) and preprocessed (dashed) daily data counts (fig. VIII 2), where the preprocessor merges different input files into one, reformats heterogeneous data formats, and screens data for timeliness. Monthly and latitude band averaged vertical profiles of rms fits to the guess (solid) and analysis (dashed) (fig. VIII 3), and sample sizes for both temperatures and winds. Note that in the tropics, below 400 hPa, the first guess is about half a degree colder than both the satellite and raob temperatures; after the analysis, the fit to the raobs is much better, but it does not change as much with respect to the satems reflecting the smaller weights those data are given. Monthly and latitude band averaged profiles of the fit to the guess and percentages of accepted and rejected temperatures and winds (fig. VIII 4), when not plotted, the rejected profiles are ouside the range. Daily latitude band averaged vertical profiles of rms fits to the guess for rawinsondes and pibals, normalized to the value at 100 hPa, a.k.a. "curtain plots" (fig. VIII 5). These plots are particularly useful to point out specific problems with rawinsondes or the first guess. For example, the effect of a "cold start" from an old operational first guess for July 1, 1985 is apparent.

IX. Pilot Reanalysis results (Mo, Wang, White, Salstein, Kalnay)

In this section, the results of several diagnostic studies performed by Kingtse C. Mo and J. Wang, Glenn White, and D. Salstein are briefly described. Together they provide a fairly good description of the relative robustness of different reanalysis fields. For these assessments we have used results from comparisons of operational and parallel assimilations comparing the impact of horizontal and vertical resolution, and of the change in convective parameterization. In addition, several Pilot analyses for July and August 1985 were performed. After a number of errors were discovered and corrected through this process (in both the Reanalysis system and in the operational system!), one final 'test' was made, wherein the operational analyses were exactly reproduced. During this process we also learned that <u>the cold start spin up/down time for variables in the tropics, unlike the mid-latitudes, is as long as two weeks</u>. Since then, three pilot runs were completed. *Pilot 2:-Control* July and August 1985, starting with July 1; *Pilot 4:* As Pilot 2 but NOSAT and *Pilot 13:* As pilot 2, but starting with July 15 1985, to test whether the analysis is sensitive to a cold start beyond two weeks.

The following two figures estimate the percentage of unexplained variance of the NOSAT analysis with respect to the SAT analysis, and as such summarize the reliability of the NOSAT analysis. Figure IX (gamma) defines the percentage of daily variability in the SAT analysis not explained by the NOSAT:

$$\gamma = < \overline{[(SAT-\overline{SAT}) - (NOSAT-\overline{NOSAT})]^2} > / VAR$$

where the bar represents a monthly average over the 31 days of August, and the daily variance (denominator) is computed as the numerator but for the SAT term only. We find that there is no significant difference in the NH SAT and NOSAT analysis even on a daily basis, with γ well below 10% North of 20N. In the tropics, were the daily variability itself is quite small, the NOSAT captures between 90% and 60% of the variance in the daily tropical variability. In the Southern Hemisphere, the daily variations are remarkably well captured in mid-latitudes: between 20S and 60S the explained variance by the NOSAT is between 80% and 90% of the variance itself. The largest differences can be found in the SH south of 60°S, and over Antarctica, the ratio becomes larger than one. The larger ratios in the upper levels at high southern latitudes, up to 90%, may be associated with the presence of very unstable short baroclinic waves, apparently due to the presence of a rigid lid in the model near the stratospheric winter jet. Figure IX (delta) estimates the NOSAT unexplained variance for the monthly averaged stationary eddies with respect to the SAT stationary waves:

$$\delta = \sum [(\overline{SAT}* - \overline{NOSAT}*)^2] / VAR(\overline{SAT}*)$$

where the sum is an average over longitude, the star is the deviation from the zonal mean and the variance is again a similar sum but computed for the SAT term only. Fig. IX (delta) shows that the NOSAT has captured well over 90% of the stationary waves variance for August 1985 in the NH, and in the SH midlatitudes. These encouraging results indicate that it is indeed worth performing a reanalysis for the years before satellite data was available, because a large component of both the daily and monthly anomalies can be captured even in the absence of the satellite data.

a) Impact of satellite data and resolution (Mo and Wang, and Salstein)

Impact of the initial guess (cold start) and of satellite data

To determine the impact of satellite data on the CDAS/Reanalysis System, we have performed diagnostics on a pair of data assimilation runs using the forecast model at the horizontal resolution T62 for the months of July and August 1985 (Pilot experiments 2 and 4), run with the 28 level model. There were two SAT runs: one experiment started at July 1 and another one started at July 15, in order to test the extent to which the first guess influences the analysis once the spin up period is over. For the August monthly mean climate variables, there is no significant difference between two Sat runs. We then started a run from July 1 where the satellite data were not included. This run is called NOSAT run. Because of the spin up problem associated with an initial guess made in 1985 (which the previous experiment showed is over after two weeks), the diagnostics will be presented for August only. The climate variables calculated on the basis of Sat and Nosat runs were compared and many variables were also compared to the old NMC analysis from our Climate Diagnostics Data Base (CDDB).

In addition to the primary variables including winds, temperature, heights and humidity, we also examined second order products such as fluxes and eddy transports. We studied the rainfall from the 0-6 hours forecast cycle for both SAT and NOSAT runs and compared it with the satellite derived product such as the GPI in the tropics and the Climate Division monthly mean precipitation data in the United States.

Primary variables

Fig. IX 1a shows the zonal mean cross section in the latitude pressure plane of the temperature difference between the NOSAT and the SAT analyses. The NOSAT analysis shows lower temperatures in the tropics and higher temperatures in the Southern Hemisphere middle to high latitudes above the 200-mb level. This may imply that the satellite data help to resolve the position of the tropopause better. The horizontal distribution of the monthly mean temperature differences at 200 hPa and 100 hPa can be seen in Figs. IX 1b and 1c respectively. The relative coldness at 100mb for NOSAT evenly distributed in the tropics. The warming at 200 mb level is concentrated at the Antarctica, with negligible differences elsewhere.

Fig. IX 1c shows the zonal mean cross section in the latitude-pressure plane of the v wind difference between SAT and NOSAT. The Hadley circulation represented by the zonal mean meridional wind in the tropics for NOSAT is up to 0.5m/s weaker than SAT, and extends to the 50 mb level. The major difference at 250 mb (Fig.IX 1d) occurs in the eastern Pacific near South America, where data coverage is sparse. Differences north of 10° N are very small. In the southern oceans, the wind in the Indian ocean is up to 3m/s weaker. The difference of meridional wind at 250 mb between NOSAT and SAT is only slightly larger than the difference between the SAT and the original operational analysis (CDDB) (Fig. IX 1f). At the 850 mb level, the differences in meridional wind (Fig.IX 1g) again appear in the eastern Pacific, in the southern oceans and in the Indian ocean. The cross equatorial flow for NOSAT is stronger than that for SAT. The differences in the zonal velocity between the SAT and NOSAT analyses (not shown) are generally small.

For heights, the major differences are confined to the Antarctica and the pattern is consistent with the temperature difference. To show an example, we plotted the 500 mb difference between SAT and NOSAT in Fig. IX 2a. The standard deviation of daily heights with respect of the monthly mean for August averaged over six years (1987-1992) based on the NMC operational analyses is also given in Fig.IX 2b. There is no noticeable difference in the Northern Hemisphere between SAT and NOSAT. There are differences in the southern oceans, but smaller than the standard deviation and also smaller than the differences between the SAT and the old analyses from CDDB (not shown).

Eddy statistics

The largest relative differences between SAT and NOSAT are in eddy statistics, and they are also confined mostly to the Southern Hemisphere. Figs. IX 2c and d show the eddy momentum transport for SAT and the differences between SAT and NOSAT. For the heat transport (Figs. IX 2e and f), the differences are confined to the Southern Hemisphere, and to thetropics south of 10°N. In general, the poleward heat transport is weaker for the NOSAT run except at the 850 mb level.

Rainfall

The global mean rainfall from the 0-6 hour forecasts averaged over August 85 for SAT is 2.69 mm/day and for NOSAT is 2.88 mm/day, a difference of only 0.19 mm/day. Figs. IX 6a and b show the estimated rain in the tropics and in the United States respectively. They can be compared to the satellite observations from the GPCP and the Climate Division rain fall data over the United States (Fig IX 6c). There are topographically related systematic errors such as somewhat excessive rain over islands, but this effect is far smaller than it was before the new cumulus parameterization was introduced (see also next subsection). The differences between the SAT and NOSAT are once again confined to the Southern Hemisphere, and even there they are relatively small.

When the reanalysis precipitation is compared to the satellite estimates, there some differences, e.g, the rainfall pattern in the Amazon is more broken than in the satellite data, and the precipitation over the west coast of India is stronger, but it is not clear in these cases whether the GPI is sufficiently reliable. Overall, however, the pattern pattern distribution of the tropical precipitation in the analysis is very realistic.

Over the United States the agreement with staion data is not as good. There is excessive rain in the Gulf and in Florida and some dryness over the continent (although it should be noted that the minimum contour is 25 mm in the observations and 5 mm0 in the analysis). This may be related to the model resolution, which may not resolve the topographically related features, or to the initialization of the soil moisture. The differences between SAT and NOSAT are still small.

Impact of resolution

Since we are going to use horizontal resolution of T62 for the CDAS, it is important to determine the differences between climate signals affected by the resolution. We have compared the CDAS with T62 resolution and the operational analyses (GDAS) of T126 for three months from Feb 92 to May 92. In additional to our in house study, Drs. Salstein and Black of Atmospheric and Environmental Research Inc. (AER) had compared the GDAS and the CDAS with the ECMWF analyses for May 1992. We sum up results below:

Primary variables such as winds, temperatures and heights:

The differences between GDAS (T126) and CDAS (T62) are small. The major differences can be found in the Southern Hemisphere at 100-200 mb levels. Fig IX 4 shows the May mean temperature at 100 mb for the GDAS and CDAS respectively and their difference. The largest differences are in the southern oceans where the data are sparse. The height differences are consistent with the temperatures.

Eddy and second moments statistics:

The differences are small, but they do differ. The largest differences appear across the lower stratosphere and largest values can be found over the high latitudes in the southern oceans and area just over the Antarctic. An example is given in Fig.IX 5 which plots the meridional heat flux by transient eddies. The differences between GDAS and CDAS are smaller than differences between GDAS and ECMWF operational analyses.

Hydrological cycle:

Since the moisture divergence and fluxes are dependent on the orography, we expect the largest impact of the horizontal resolution on moisture flux fields. The maximum differences are about 15% in the moisture divergence field and on the 6 hour forecast of rainfall. The T126 model results in more rainfall than T62. The largest local differences are in the central/north America as indicated by E-P and D(iv) averaged over March 1992 (Fig.IX 6). The pattern for GDAS is smoother and more realistic than for CDAS. There is very little differences in the forecast evaporation.

b) Impact of the change in cumulus convection and "signal to noise" estimates (White)

The Development Division routinely tests new versions of the NMC global analysis/forecast system by running them every day in parallel to the operational system and comparing the results. Several tests with implications for reanalysis have been run. For some years a T62 version of the operational model has been run; a 28-layer version of the model with Pan-Grell convection (which became the operational global model on Aug. 10) was also tested for some months against the then-operational 18-layer model with Kuo convection (both were run at T126).

Fig. IX 7 compares the global water budget in T62 and T126 analyses and forecasts during Apr.-Sept. 1992. Both systems exhibit a small "spin-up" during the forecasts. T62 has less precipitation than T126 (especially over tropical continents), although all the values of global mean precipitation shown are within the range of climatological estimates (Legates, 1993). Fig. IX 8 compares the global water budget for a 28-layer system with Pan-Grell convection to an 18-layer system with Kuo convection during June 29-July 20, 1993. The Pan-Grell system shows less change with time than does Kuo convection.

Another method we have used to assess the effect of small differences in the analysis system was to calculate the spatial root-mean-square (rms) monthly mean difference between the analyses produced by the two systems (the noise produced by differences in the analysis systems) and compare it to the spatial rms of the monthly mean operational analysis (a measure of the climate signal). Geopotential heights, winds, vorticity, divergence, streamfunction, velocity potential and transient eddies at various levels were compared, as were precipitation, surface fluxes and top-of-the atmosphere radiative fluxes for shorter 10-day periods. Most of the fields examined were little affected by the differences in analysis systems. During January 1993 only upper level divergence, 70 mb winds, and precipitation in the tropics and Southern Hemisphere extratropics showed rms time-mean differences greater than 50% of the spatial rms of the time-mean operational analysis.

Results for a mid June-mid July 1993 test of the 28-layer system with the Pan-Grell convection against the 18-layer system with Kuo convection were similar. The spatial rms of the time-mean differences were greater than 50% of the spatial rms of the time-mean operational analyses only for upper-level divergence, 150 mb meridional wind (only in the tropics), 70 mb winds, precipitation, cloudiness and radiation. Differences in the last two reflected an increase in cloudiness with Pan-Grell, partially correcting a bias towards too little cloudiness in the NMC operational global analysis/forecast system. In general differences were largest in the tropics, reflecting a lack of data compared to the Northern Hemisphere, a lack of usefulness of satellite temperature soundings compared to the Southern Hemisphere and the greater direct influence of poorly parameterized diabatic heating on the tropical atmosphere than in the extratropics. Changes in the ECMWF analysis/forecast sytem have also had the greatest impact on fields related to the divergent circulation in the tropics (Trenberth and Olson, 1988 ; Hoskins et al., 1989).

Fig. IX 9 shows the time-mean 150 mb divergence for June 13-July 12, 1993 from (top) the 28-layer analysis system with Pan-Grell convection and (middle) the 18-layer analysis system with Kuo convection and (bottom) the difference between them. The spatial rms of the difference (bottom) is 96% of the spatial rms of the 18-layer analysis (middle) between 20S and 20N, 90% between 20S and 80S and 75% between 20N and 80N, implying that noise due to analysis differences is as large as the climatic signal. However, an examination of Fig. reveals very similar large-scale patterns in the two analyses, although large differences in magnitudes exist on smaller scales. The same examination of 150 mb velocity potential found that the ratio of the rms of the time-mean difference to the spatial rms of the analyzed time-mean velocity potential was only 25%, implying that the two systems do produce similar large-scale divergent circulations in the tropics and differ largely on smaller scales.

Fig. IX 10 shows precipitation during July 11-20, 1993 from (top) the 28-layer Pan-Grell analysis system, (middle) the 18-layer Kuo analysis system and (bottom) the difference between the two. The rms of the difference (bottom) is 54% of the spatial rms of the Kuo analysis (middle) between 20N and 20S. However, the two precipitation patterns are quite similar. Precipitation in both analysis systems show quite similar tropical features to satellite observations of outgoing longwave radiation.

Upper level tropical divergent flow, precipitation and stratospheric winds appear to be most sensitive to small changes in the NMC analysis system. The large-scale pattern of upper-level divergent flow in the tropics appears to be fairly robust; however, the magnitude and smaller-scale features of the upper-level divergent flow is still poorly defined by a modern state-of-the-art analysis system.

Fig. IX 11 displays the actual rms differences between monthly mean 500 mb height in the operational NMC global analyses, analyses from the T62 and Pan-Grell parallel tests, and analyses from other operational NWP centers for Jan. and July 1993. In the Northern Hemisphere extratropics, 500 mb rms height differences between the NMC operational analyses and the parallel tests are slightly less than 6 meters, suggesting that the uncertainty in monthly mean 500 mb height is 6 m in the Northern Hemisphere. The monthly mean difference between NMC analyses and other centers is slightly larger, reflecting in large part a larger area mean difference. The uncertainty in the Southern Hemisphere is slightly larger.

Fig. IX 12 displays rms differences between monthly mean heights at various levels in the operational NMC global analyses and in T62, Pan-Grell and UKMO analyses. T62 and Pan-Grell analyses display an rms difference of less than 4 m at 850 mb; UKMO analyses display a considerably larger difference, reflecting largely differences under orography. RMS differences increase with height and are larger in the Southern Hemisphere. Differences between the NMC and UKMO operational analyses exceed the differences between the NMC operational and experimental analyses in the Northern Hemisphere except at 100 mb. However, the RMS difference between T62 and T126 NMC analyses exceed other differences in the Southern Hemisphere.

Fig. IX 13 displays rms differences in monthly mean 850 and 200 mb winds from the operational NMC global analyses, experimental NMC analyses and UKMO analyses in the extropical Northern and Southern Hemisphere and the tropics. Differences are least in the Northern Hemisphere. RMS differences between NMC operational and experimental analyses are 0.8 m/s or less at 850 mb and 1.8 m/s or less at 200 mb.

These figures indicate the precision with which modern analysis systems can determine monthly mean meteorological fields and can be regarded as a lower estimate of the accuracy with which such fields can be determined. Since all the analyses used employ very similar data bases, it is likely that the true error is larger than the differences between the different analyses, since errors due to data gaps and measurement errors would be similar in the different analysis systems.

Fig.IX 1 (a) Zonal mean temperature difference between NOSAT and SAT for August 1985. Contour interval 1 C.
(b) Temperature difference between SAT and NOSAT for August 1985 at 100 mb . Contour interval 1C.
(c) Same as (b), but at 200 mb. Contour interval 2C.
(d) Zonal mean meridional wind difference between NOSAT and SAT for August 1985. Contour interval 0.1 m/s
(e) V-Wind difference for August 1985 between NOSAT and SAT at 250 mb. Contour interval 1 m/s. zero contour is omitted.
(f) V-wind difference for August 1985 between the NMC analyses and SAT. Contour interval 2 m/s. Zero contour is omitted.
(g) Same as (b), bit for the 850 mb V wind. Contour interval 0.5 m/s and zero contour is omitted.
Fig. IX 2 (a)500 mb height diffference for August 1985 between NOSAT and SAT. Contour interval 20 m .
(b) Standard deviation for 500 mb heights for August from the CDDB.
Contour interval 10 m.

Fig. IX 2 (cont.): Zonal mean of meridional transport of momentum by transient eddies (U'V') for August 1985 (c) for SAT run contour interval 5 $(m/s)^2$ and (d) difference between NOSAT and SAT. contour interval 2.5 $(m/s)^2$.
Fig. IX 2 (cont.): Zonal mean of meridional heat transport by transient eddies for August 1985. (e) for SAT run. contour interval 2 (m/s)*K;(f) difference between NOSAT and SAT . Contour interval 1 (m/s)*K.
Fig. IX 3 Rain from 0-6 hours forecast in the tropics for (a) SAT and (b) NOSAT , contour 2.5 mm/day . (c)rainfall accumulation for August in the United States for SAT. Contour interval 50,75,100,200 300 400 mm , (d) same as (c), but for NOSAT.
(e) rainfall for August from GPI. Contour interval 2.5 mm/day and (f) rain accumulation for August 1985 from the climate division data. contour interval 25 ,50, 75, 100, 150, 200, 300, 400 mm.
Fig. IX 7 Global mean precipitation and evaporation as a function of forecast length in NMC T126 and T62 analyses and forecasts during April - September 1992.
Fig. IX 8 Global mean precipitation and evaporation as a function of forecast length in NMC analyses and forecasts from an 18-layer system with Kuo convection, and 28-layer system with Pan-Grell convection during June 29 - July 20, 1993.
Fig. IX 9 Time averaged 150 mb horizontal divergence for June 13 -July 12, 1993 from NMC analyses with (top) 28 layers and Pan-Grell convection and (middle) 18 layers and Kuo convection. (Bottom) The difference (top minus middle). Contour interval 2x10 -6/sec. Positive values shaded in top and middle; negative values shaded in bottom.
Fig. IX 10 Precipitation for July 11-20, 1993 from NMC 0-6 hr forecasts with (top) 28 sigma layers/Pan-Grell convection and (middle) 18 sigma levels/Kuo convection. Contour interval 4mm/day, beginning at 1mm/day shaded. (Bottom) The difference (top minus middle). Contour interval 4mm/day, with contours at -2 and 2mm/day. Values less than -2mm/day shaded.
Fig. IX 11 Spatial RMS differences in 500 mb monthly mean height between NMC operational analyses and NMC experimental analyses and analyses by other NWP centers during January and July 1993. T62 refers to NMC experimental analyses at T62 horizontal resolution, PG to NMC experimental analyses with 28 sigma layers and Pan-Grell convection, EC to ECMWF, UK to UKMO, FN to FNOC and JM to JMA.
Fig. IX 12. Spatial RMS difference in monthly mean height at different pressure levels between NMC experimental analyses and UKMO analyses during January and July 1993.
Fig. IX 13 As Fig. 12, except for winds at 850 and 200 mb.

X. Coupling with ocean data assimilation (Leetmaa and Reynolds)

For the first phase of CDAS/Reanalysis it has been decided that the atmospheric analysis will be coupled with the SST reanalysis of Reynolds (see section a), from 1982 onwards. For the earlier periods we will use the Global sea-Ice and SST Analysis (GISST) that the UKMO has kindly offered to make available (David Parker, pers. comm., Parker et al, 1993). In addition, and ocean 4-D reanalysis will be also performed, but with a one-way interaction in the coupling with the atmospheric reanalysis (see section b).

a) The NMC SST Analysis (Reynolds)

The U.S. National Meteorological Center (NMC) routinely produces a 1° gridded SST analysis using optimum interpolation (OI) (e.g. see Gandin, 1966; and Thiébaux and Pedder, 1987). The analysis is produced both daily and weekly following the method of Lorenc (1981). Both versions use seven days of *in situ* (ship and buoy) and bias corrected satellite SST data as well as SSTs estimated from sea ice coverage.

The analyses are determined relative to a first-guess or predicted analysis. The first-guess field is always the preceding analysis. Other investigators have used climatology as a first guess. Because time scales of SST anomalies have been found to be of the order of months, the analysis from the previous week is a better forecast of SST than climatology.

The bias correction is carried out to correct any large-scale satellite biases. This correction is necessary because of the strong satellite biases have occurred (Reynolds, 1993) and because the OI method assumes that the data are unbiased. This adjustment uses the Poisson technique of Reynolds (1988) and Reynolds and Marsico (1993) to provide a smooth correction field. In this method, regions with sufficient *in situ* observations become internal boundary conditions and regions that are ice-covered become external boundary conditions. The SSTs for the remaining grid points are determined by solving Poisson's equation where the forcing term is defined as the Laplacian of the satellite analysis. This method adjusts any large scale satellite biases and gradients relative to the boundary conditions defined by the *in situ* data.

The OI is only optimal when the correlations and variances are known for the analysis increment and each type of data increment. These statistics have been evaluated to show that ship SST observations have larger errors (1.3°C) compared to the errors of buoy and satellite SSTs (0.3-0.5°C). In addition the e-folding correlation scales have been found to range between 500 and 1200 km. Complete documentation for the OI with the satellite bias correction has been conditionally accepted for publication (Reynolds and Smith, 1993).

An example of the OI SST analysis is shown in Fig. 1 for the week, 13 to 19 January 1991. This can be contrasted with the older blended analysis of Reynolds (1988) and Reynolds and Marsico (1993), which is shown in Fig. 2 for a 15 day period, 9 to 23 January 1991. The time period for the blended analysis was reduced to the practical minimum for that scheme and chosen so that the centers of the two analysis periods coincide. The higher resolution of the OI analysis is apparent. In the tropics, the equatorial eastern Pacific and Atlantic cold tongues are more realistically shown in the OI. At higher latitudes, the OI shows tighter gradients, particularly in the Gulf Stream, the Kuroshio and the Malvinas (Falklands) current regions. This higher resolution depiction qualitatively agrees with patterns shown in the satellite data and the patterns simulated by high-resolution ocean models.

The daily version of the OI was operationally implemented at NMC beginning on February 20, 1991 and became the SST boundary condition for the NMC global models. Because the data from any one day would enter the analysis seven times during one week, the OI error statistics where adjusted to give the data the same weight in the daily and weekly versions.

The weekly version of the OI analysis has been computed for the period January 1985 to present. These fields are now used at NMC for reanalysis efforts and are available to other users via the NMC Internet file server. The additional period December 1981 to December 1984 will be added before the end of the year. It is not practical to extend the period prior to December 1981, because the present operational satellite instrument first became operational in November 1981.

The OI analysis with the bias correction is still evolving. Once the OI has been analyzed from December 1981 to present, the OI statistics will be reevaluated using the entire data set. The present version of the OI with the bias correction is a significant improvement over the earlier NMC analysis and over any other analysis that uses uncorrected satellite data.

b) The ocean analysis system (Leetmaa)

At NMC, the Coupled Model Project has been performing 4-D ocean analyses for a number of years in order to document more thoroughly current and past climate variability. Furthermore, these analyses served as the initial conditions and verification fields for the coupled ocean-atmosphere model used for multiseason forecasting (Ji, et al., 1993). Since most of the potential extended predictability is thought to be the result of coupled interactions in the tropics, the focus at NMC has been the development of the analysis capability in this region.

Routine weekly ocean model based analyses are performed for the Pacific and Atlantic Basins. Reanalyses have been finished for the Pacific basin for the period June 1982 to December 1992. A reanalysis for the Atlantic for the same time period is underway. The ocean model that is used was developed at the Geophysical Fluid Dynamics Laboratory (GFDL). The Pacific model has a domain that extends from 45S to 55N and 120E to 70W. The Atlantic domain extends from 100W to 20E and 50S to 65N. The bottom topography is variable and there are 28 levels in the vertical. The zonal grid spacing is 1.5 degrees in the Pacific and 1 degree in the Atlantic. The meridional grid spacing is 1/3 degree within 10 degrees of the equator and gradually increases outside this zone to 1 degree poleward of 20 degrees. Within ten degrees of the northern and southern boundaries the model fields are relaxed to climatological estimates. A Richardson number formulation controlled vertical mixing is used in the upper ocean. The lateral mixing is formulated as proportional to the square of the equivalent horizontal wave number. Our plans are to expand the analyses to a global domain in early 1994.

The formal structure of the data assimilation system was described by Derber and Rosati (1989). This a three dimensional variational technique that is applied continuously in time. Presently only thermal data is used in the analyses. All available temperature data from ships, satellite estimates, drifting and moored buoys, and expendable bathythermograph (XBT) are used. Extensive quality control procedures have been developed to screen the data before they are used. Corrections are made to the model thermal fields in the upper ocean down to 720 meters. This depth range contains the maximum depth of the bulk of the available subsurface thermal data from the T4 and T7 type XBT probes. Surface observations are kept in the analyses for two weeks and subsurface observations are kept in for four weeks with weights varying linearly in time. Maximum weight is given at the observation time with minimum weights at the beginning and end of the time interval the data is being used.

Routinely, two weeks after the fact, the Atlantic and Pacific models are run for a week. The time delay is required in order to allow the XBT data to be kept in the assimilation for the full four weeks. The models are forced with a weekly averaged stress field derived from the 4 times per day near surface winds produced by NMC's global atmospheric analyses. These winds are converted to stress using a constant drag coefficient of $1.3*10^{-3}$. The net heat flux used to force the model is set to zero in order to facilitate the evaluation of the heat fluxes in the analysis system. The net fresh water flux is set to zero; once a year the salinity field is restored to the mean climatological field of Levitus (1988). A fuller description of the analysis system is currently being written.

Using this system, ocean reanalyses have been or are currently being performed for both basins. For this purpose all available historical subsurface data was obtained from the archives and edited for the time period June 1982 to the end of 1992. These were merged with the data that was available in realtime. (A global sea surface temperature reanalysis is also in progress for this time period. See section IX a.).

Two reanalyses for this period have been finished for the Pacific. These use slightly different versions for the stress forcing, additional data was added, and further quality control was performed. The Atlantic reanalysis will be finished in the coming months. These data sets are being used to document the coupled ocean-atmosphere variability that took place this past decade, and were used to develop the coupled forecast model that is currently being used for experimental multi-season forecasts at NMC.

Work on implementing a global model-based ocean anaysis system will start in early 1994. The routine weekly analysis capability will be implemented first; the global reanalysis for the period 1982 to present will be started in late 1994, when the atmospheric reanalysis becomes available, and should take a few months for completion. With respect to earlier periods, the scarcity of subsurface ocean data implies that a meaningful reanalysis can only be done for the Northern Hemisphere, and only for the period starting in the late 1960's. Nevertheless, once a consistent set of forcing fields is available from the atmospheric reanalysis, we plan to perform the ocean reanalysis for the whole period (1958-present).

Figure Captions

Fig. X.1. Mean SST from the OI using the assumed statistics for the week: 13-19 January 1991. The contour interval is 1° with heavy contours every 5° starting with 0. The -1°C contour is dashed; SSTs below -1.79°C are shaded.

Fig.X. 2 Mean SST from the blended analysis for the 15 day period: 9-23 January 1991. Otherwise as in Fig. 2.

XI. Distribution (Kalnay)

The question of the distribution of the CDAS/Reanalysis archives is of paramount importance, but has not yet been resolved, and should be discussed with NOAA OGP management, the Advisory Committee, and others. During the April 1991 NMC/NCAR Workshop on Reanalysis, it was repeatedly stressed that the availability of user friendly distribution was essential to the success of the project (otherwise, as somebody noted, it would be a "write only" data set).

Several factors should be taken into account during the final discussions on data distribution:

a) During the 1991 Workshop, M. Blackmon offered that the NOAA/ERL/Climate Research Division could act as "Library of Congress" for the Reanalysis data and distribution. He has since estimated that the hardware and software needed for this job would require considerable funding.

b) Copies of the CDAS/Reanalysis output will be sent to both NOAA/NCDC and NCAR.

c) There are going to be several archives of different sizes: a) The main BUFR data and GRIB analysis, diagnostics, sigma and pressure files, etc will take about 10-20 STK (0.2GB each) cartridges per month. b) The condensed "time series" archive, which should fit into one cartridge per month. In addition, two more special archives may be created: c) The forecast archive (if the 5-day forecasts every 5-10 days are carried out), and d) the "North American" GRIB archive, with the fields most probably useful for NWS Forecast Offices, the creation of a Perfect Prog, etc. The latter two archives will, hopefully, also fit into one cartridge per month.

d) Of the reduced archives (one cartridge, 0.2GB per month), it may be possible and desirable to create CD-ROM (0.4 GB each) or Exabyte (4GB each) data sets for wider distribution.

e) NMC has implemented a File Server accessible by Anonymous FTP. Dr. McPherson has expressed his wish that some of the CDAS/Reanalysis products be available also in this archive. It may be possible to design an appropriate subset of the data (e.g., the last month of CDAS, and the latest year of reanalysis completed, at a few pressure levels) that can be periodically posted and refreshed for anonymous user retrieval.

f) NASA/GSFC (Siegfried Schubert, pers. comm., Schubert et al, 1993) is creating a 5 year Reanalysis, which will be available in the NASA EOSDIS Mass Storage System, designed for easy browsing and retrieval by the research community. Such a system would be also ideally suited for the purposes of CDAS/Reanalysis data distribution, but our complete data base may overwhelm that Facility. Dr. Schubert explained that Goddard has two projects (DAC0 and DAC1) for mass storage and distribution within EOSDIS. Dr. Milton Halem, Chief of the Space Data and Computing Division has also indicated a willingness to consider posting the NMC/NCAR Reanalysis in the NASA system if feasible, at least the lower volume data sets.

g) It should be explored whether the Lawrence Livermore Laboratory Atmospheric Modeling Intercomparison Project (AMIP) could also be a center for distribution.

XII. Discussion

The CDAS/Reanalysis has many applications not only for research in general, but also for NOAA's operational needs. Some of the research studies made possible by the availability of a very long reanalysis dataset are:

- ENSO simulations, providing estimates of surface fluxes for a long period
- Climate change, especially through the availability of Optimal Averaging
- Dynamical and predictability studies
- Impact of the changes of the observing systems on the forecast skill
- Budget studies, e.g., for the hydrological cycle
- Greenhouse gases transport studies (using the restart files)
- Length-of-day and angular momentum studies

NOAA's own operational needs will also be served by this data base. Some examples of possible applications are:

- CAC's monitoring of short term climate changes, by comparing the frozen CDAS with the long reanalysis
- ENSO operational predictions with the NMC coupled ocean atmosphere model, which will benefit from the available data for training and verification of the system
- The special North American reanalysis archive should be useful to the NWS Forecast Offices as a regional climatological data base
- The NWS will be able to create a very long "Perfect Prog" data base, relating the reanalysis fields to the station observations. This can be combined with Model Output Statistics (MOS) in which model forecasts are used to predict the analysis, requiring only short training. This will give a more accurate (because of the long Perfect Prog record) and more flexible MOS system to provide station significant weather forecasts
- The global reanalysis should provide a very useful data base for retraining NHC's statistically-based hurricane forecast model.
- Very importantly, many advanced components (e.g., BUFR with QC information, and the Automatic Monitoring System) developed for the CDAS/Reanalysis will be also implemented operationally into the forecasting system

In all these applications, it is important that the users be aware that not all the fields are equally reliable. It is obvious that the changes in the observing system will introduce inhomogeneities in the reanalysis. The pilot studies, the comparisons with the analyses of other centers, and the studies of the impact of horizontal resolution and cumulus parameterizations, have indicated that some fields are more reliable than others. Among the most robust products of the reanalysis are the Northern Hemisphere extratropics fields below 100 hPa. The Southern Hemisphere mid-latitudes are also reasonably well captured by the reanalysis, even in the absence of satellite data. It is important to distinguish here between the relative accuracy required for climate anomaly diagnostics, for which the reanalysis precision will probably be acceptable, and the accuracy needed for skillful medium range weather forecasts, for which analysis without satellite data are not acceptable. The least reliable products of the reanalysis are the fields above 100 hPa, and those in the high latitudes of the Southern Hemisphere, both of which show strong sensitivity to both data and resolution. Even for these fields, however, the Reanalysis may contain useful signals in the intermonthly and interannual variability.

XIII. Acknowledgements

This has been, and continues to be, the largest and most ambitious project ever undertaken at NMC (jointly with NCAR). The enthusiastic support of the Director of NMC, Ron McPherson, and of the President of UCAR, Rick Anthes have been essential in giving us with the courage to undertake it.

The Advisory Committee's role has been crucial: during the formal meetings and mail reviews the Committee made major suggestions for improvements. We are most grateful to the Chair of the Advisory Committee, Prof. Julia Nogués-Paegle who gave us not only advice and many important suggestions, but also constant support. Prof. Mark Cane originally proposed that NMC carry out a long reanalysis, in addition to the CDAS.

Many people have contributed to the project within NMC. We are most grateful to Fran Balint for making hardware resources available, Wayman Baker for his initial efforts in the organization of the project. John Derber, Yuejian Zhu, Jie Meng, Mark Iredell, Laurie Morone, Dennis Keyser, Robert Grumbine Hua-lu Pan, and many others scientists from the Development Division for their many technical contributions and advice.

NESDIS personnel have been most helpful in providing a data base of TOVS operational data. Brian Doty from COLA, with the encouragement of Drs. J. Kinter and J. Shukla, generously provided the GrADS graphics system, and developed considerable new components required by the Reanalysis. David Parker and Chris Folland's offer to make available the UKMO GISST reanalysis of ice and sea surface temperature for the period earlier than 1982 has been gratefully accepted. The Japanese Meteorological Office has provided us with cloud track wind data, and offered additional observations not available on GTS. ECMWF has offered to provide data for those periods where the NMC/NCAR data base is not available.

This project would have been impossible without the support of the NOAA Office of Global Programs. We are in particular most grateful to Michael Coughlan, the Program Manager, who has been helpful and knowledgeable, and made the proposal and budget submission and approval process seem almost easy. Ken Mooney support and suggestions were also very helpful.

Similarly, the support of the NCAR effort by the National Science Foundation has been essential, and in particular we want to acknowledge the support of Jay Fein.

XIV. References

Chelliah, M.: Early results from the Pilot Project of the NMC/NCAR CDAS/Reanalysis Project. Proc. Clim. Diagn. Workshop, Boulder, Co 1-5 Nov 1993.

Collins, W.G. and L.S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. *Mon. Wea. Rev.*, **18**, 2754-2767.

Derber, J.D., and A. Rosatti, 1989. A global oceanic data assimilation system. J. Phys. Oceanogr., 19, 1333-1347.

Derber, J. C., D. F., Parrish and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. Wea. and Forecasting, , **6** 538-547.

Gandin, L.S., 1993: Optimal Averaging of Meteorological Fields. *NMC Office Note 397.*

Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields.* In Russian, Gidrometeor. Isdat., Leningrad. [English translation, Israeli Program for Scientific Translations. Jerusalem, 1966, 242 pp.]

Gandin, L.S.: 1988: Complex quality control of meteorological observations. *Mon. Wea. Rev.*, **116**, 1138-1156.

Hoskins, B.J., H.H. Hsu, I.N. James, M. Masutani, P.D. Sardeshmukh and G.H. White, 1989: Diagnostics of the Global Atmospheric Circulation Based on ECMWIZ Analyses 1979-1989. World Meteorological Organization, Geneva, Switzerland, WMO/TD No. 236, 217 pp.Res., Boulder, CO. NCAR/TN-301+STR, 81pp.

IPCC (1990): Climate Change: The IPCC Scientific Assessment. WMO/UNEP International Panel on Climate Change.

Jenne, R. 1993: Various documents describing the data base (see section III).

Ji, Ming, A. Kumar, A. Leetmaa, 1993. Development of a seasonal climate forecast system using coupled ocean-atmosphere model at the National Meteorological Center. Submitted to the Bulletin of the AMS.

Kalnay E., and R. Jenne, 1991: Summary of the NMC/NCAR Reanalysis Workshop of April 1991. American Meteorological Society, **72**, 1897-1904.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. Wea. and Forecasting, **4**, 334-342.

Kanamitsu, M., J.C. Alpert, K.A. Campana, P.M. Caplan, D.G. Deaven, M. Iredell, B. Katz, H.L. Pan, J. Sela and G.H. White, 1991: Recent changes implemented into the global forecast system at NMC. NMC Notes, 0001-0012.

Kistler, R. E. Kalnay and M Kanamitsu, 1993: The NMC CDAS/Reanalysis Project: A status report. Proc. Clim. Diagn. Workshop, Boulder, Nov 1-5 1993.

Legates, D.R., 1993: Global Observations, Analyses and Simulation of Precipitation, WGNE/GEWEX Workshop, NMC, Camp Springs, MD, USA, 27-30 Oct. 1992. World Meteorological Organization, Geneva, Switzerland, WMO/TD No. S44, 23-29.

Levitus, S. 1982. Climatological Atlas of the World Ocean. DOC, NOAA Professional Paper 13.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, 109, 701-721.

Mo, K.C., and X. L. Wang, 1993: Impact of satellite data on the CDAS. Proc. Clim. Diagn. Workshop, Boulder Co, 1-5 Nov 1993.

Parker, D. E. , C. K. Folland, M.N. Ward, K. Maskell, A. Bevan, and M. Jackson, 1993: Ocean-atmosphere climatic fluctuations on interannual to century time scales. Publ. of Hadley Centre for Climate Prediction and Research, UKMO, Bracknell, Berks., England.

Parrish, D. F. and J.C. Derber, 1992: The National Meteological Center's spectral statistical interpolation analysis system. Mon. Wea.Rev.,**120**, 1747-1763.

Reynolds, R. W., 1988: A real-time global sea surface temperature analysis. *J. Climate,* 1, 75-86.

Reynolds, R. W., 1993: Impact of Mount Pinatubo Aerosols on Satellite-Derived Sea Surface Temperatures. *J. Climate*, 6, 768-774.

Reynolds, R. W. and D. C. Marsico, 1993: An improved real-time global sea surface temperature analysis. *J. Climate*, 6, 768-774.

Reynolds, R. W. and T. M. Smith, 1993: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate* (conditionally accepted).

Saha, S., and M. Chelliah, 1993: Automatic Monitoring system for the atmospheric Reanalysis project at NMC. Proc. of Clim. Diagn. Workshop, Boulder, CO 1-5 Nov 1993

Schubert, S., R. Rood and J. Pfaendtner, 1993: An update on NASA Reanalysis effort. Subm. to BAMS.

Thiébaux, H. J., and M. A. Pedder, 1987: *Spatial Objective Analysis: With Applications in Atmospheric Science*. Academic Press, 299 pp.

Trenberth, K. E., and J. G. Olson, 1988: Intercompanion of NMC and ECMWF Global Analyses: 1980-1986. National Center for Atmos. Research

Woollen, J. S., 1991: New NMC operational OI quality control. Ninth Conference On Numerical Weather Prediction, American Meteorological Society, 24-27.

Appendix A: Participants of the NCAR/NMC CDAS/Reanalysis Project and their main areas of responsibility:

NMC/Development Division:

Eugenia Kalnay: Co-PI (with Roy Jenne)
Robert Kistler: Project Manager, design and implementation of the CDAS/Reanalysis system.
Masao Kanamitsu: Project Scientist; satellite data QC, overall design and guidance.
Jack Woollen: Design of the new BUFR data base, BUFR interface, OI QC, QC monitoring. Conversion of CQC to UNIX.
Suranjana Saha: Design of Automatic Monitoring System of output (with M. Chelliah) Climatological QC of data within the data preprocessor module.
Jie Meng: Satellite QC monitoring
Glenn White: Diagnostics and Monitoring of the system performance.
John Derber: Spectral Statistical Interpolation and adaption to CDAS/Reanalysis.
Dennis Deaven: Implementation of Optimal Averaging
Lev Gandin: Complex QC and Optimal Averaging
William Collins: Complex QC system, addition of time interpolation checks.
Laurie Morone: Complex QC of the winds.
Yuejian Zhu: Archives and data sets.
Dennis Keyser: Archive questions
Pat Astill: Graphics support (since July 1993)
Mike Fiorino: Diagnostics and graphics support (since September 1993)
Robert Grumbine: Sea Ice data

NMC/Climate Analysis Center:

Muthuvel Chelliah: Design of the Automatic Monitoring System of output (with S. Saha), minutes of the meetings, diagnostic assessment.
Chet Ropelewski: Consultation, minutes of the meetings, data base and CDAS design.
John Janowiak: Design of the time series archive (with Chelliah, Saha and Ropelewski)
Kingtse C. Mo: Diagnostic assessment of the quality of the system, impact of horizontal resolution, satellite impact, etc.
Julian Wang: Diagnostic assessments of the system
Vern Kousky: Consultation, minutes of the meetings.

NMC/Automation Division:

Fran Balint: Hardware support (Silo, Cray-EL)
Sarah Roy:
Cliff Dey:
William Cavanaugh: Office Note 29 to BUFR converters

NMC/Coupled Model Project

Dick Reynolds: SST reanalysis
Ants Leetmaa: Coupled ocean model reanalysis

NCAR

Roy Jenne: Co-PI; responsibility for the data
Dennis Joseph, Wilbur Spangler, Gregg Walters, and Joey Comeaux: Upper air data

Steve Worley and others: COADS ship data
Roy Barnes and Ilana Stern: Surface synoptic data
Dennis Joseph and Chi-fan Shih: satellite data
Wilbur Spangler: management of huge sort programs

Note: About 3 FTE are being used on Reanalysis at NCAR, plus operations support

NOAA/ERL Climate Monitoring and Diagnostics Laboratory

Scott Woodruff, Sandra Lubker and Mei Yuen: key participants in COADS

NOAA/National Climatic Data Center

Joe Elms: participant in COADS project
Dick Davis: History and status of old data sets, with Jenne

University of Missouri

Ernest Kung: Retrieval of currently unavailable data sets (e.g., China, Russia), and QC

GFDL

Kiku Miyakoda: Ocean modeling and data assimilation
John Lanzante: Fix problems with old data
Abraham Oort: Consultation, Advisory Committee

CDAS/Reanalysis Advisory Committee

First meeting, 1990

Chair: Julia Nogués-Paegle

Roy Jenne (ex officio since 1991)
Maurice Blackmon (ex officio since 1991)
Abraham Oort
Don Johnson
Mark Cane
Tom Schlatter (until 1993)
Tony Hollingsworth (since 1993)
J. Shukla (since 1993)

Appendix B

Detailed content of the main output files

All files are written 4 times a day at 00, 06, 12 and 18 GMT

1. Pressure file (analysis and 6 hour forecast guess on 2.5 degree latitude-longitude grid)

Geopotential height (gpm) at 16 levels
u-wind at 16 levels (m/s)
v-wind at 16 levels (m/s)
Relative humidity at lowest 7 levels (%)
Temperature (thermodynamic not virtual) at 16 levels (K)
Pressure vertical velocity at lowest 11 levels (Pa/s)
Surface pressure (Pa)
Surface skin temperature (K)
Pressure vertical velocity at the surface (Pa/s)
Relative humidity at the surface (%)
Temperature at tropopause level (K)
Pressure at tropopause level (Pa)
u wind at tropopause level (m/s)
v wind at tropopause level (m/s)
Vertical speed shear at tropopause level (/s)
Best (4-layer) lifted index (K)
Pressure at maximum wind level (Pa)
u wind at maximum wind level (m/s)
v wind at maximum wind level (m/s)
Surface geopotential height (gpm)
u wind at the lowest model level (m/s)
v wind at the lowest model level (m/s)
Potential temperature at the lowest model level (K)
Layer mean relative humidity (sig=1-0.40, 0.93-0.66, 0.66-0.40) (%)
Precipitable water (kg/m^2)

Standard pressure levels: 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30 20,10 hPa
All parameters are instantaneous values

2. <u>Surface flux file</u> (On the model Gaussian grid 192 x 64)

Zonal wind stress (N/m^2) (A)
Meridional wind stress (N/m^2) (A)
Sensible heat flux (W/m^2) (A)
Latent heat flux (W/m^2) (A)
Skin temperature (K) (I)
Soil moisture content (kg/m^2) (I)
Water equivalent of accumulated snow depth (kg/m^2) (I)
Downward short wave radiation flux at the top (W/m^2) (A)
Upward short wave radiation flux at the top (W/m^2) (A)
Upward long wave radiation flux at the top (W/m^2) (A)
Downward long wave radiation flux at the surface (W/m^2) (A)
Upward long wave radiation flux at the surface (W/m^2) (A)
Downward short wave radiation flux at the surface (W/m^2) (A)
Upward short wave radiation flux at the surface (W/m^2) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
High cloud top temperature (K) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Middle cloud top temperature (K) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Low cloud top temperature (K) (A)
6 hr total precipitation (kg/m^2) (Acc)
6 hr convective precipitation (kg/m^2) (Acc)
Ground heat flux (W/m^2) (A)
Land-sea mask (1=land; 0=sea; 2=sea-ice) (I)
u wind at 10 m (m/s) (I)
v wind at 10 m (m/s) (I)
Temperature at 2 m (K) (I)
Specific humidity at the lowest model level (kg/kg) (I)
Surface pressure (Pa) (I)

(A) .. Average in 6 hr forecast
(I) ... Instantaneous value at forecast hour 6
(Abc) .. Accumulation in 6 hour forecast

3. Diagnostic file at FT=6 hr (on the model Gaussian grid 192 x 94)

Surface skin temperature (K) (I)
Soil moisture content (kg/m^2) (I)
Water equivalent of accumulated snow depth (kg/m^2) (I)
Soil temperature at 10, 50 and 500 cm (K) (I)
Surface roughness (m) (I)
Total 6 hour accumulated precipitation (kg/m^2) (Acc)
Convective 6 hr accumulated precipitation (kg/m^2) (Acc)
Sensible heat flux (W/m^2) (A)
Zonal wind stress (N/m^2) (A)
Meridional wind stress (N/m^2) (A)
Latent heat flux (W/m^2) (A)
Albedo (%) (I) *
Downward short wave radiation flux at the top (W/m^2) (A)
Upward short wave radiation flux at the top (W/m^2) (A)
Upward long wave radiation flux at the top (W/m^2) (A)
Downward long wave radiation flux at the surface (W/m^2) (A)
Upward long wave radiation flux at the surface (W/m^2) (A)
Downward short wave radiation flux at the surface (W/m^2) (A)
Upward short wave radiation flux at the surface (W/m^2) (A)
Total high level cloud cover (%) (A)
Nearby model level of high cloud top (int) (A)
Nearby model level of high cloud bottom (int) (A)
Total middle level cloud cover (%) (A)
Nearby model level of middle level cloud top (int) (A)
Nearby model level of middle level cloud bottom (int) (A)
Total low level cloud cover (%) (A)
Nearby model level of low cloud top (int) (A)
Nearby model level of low cloud bottom (int) (A)
Convective cloud cover (%) (A)
Visible beam downward solar flux at the surface (W/m^2) (A)
Visible diffuse downward solar flux at the surface (W/m^2) (A)
Near IR beam downward solar flux at the surface (W/m^2) (A)
Near IR diffuse downward solar flux at the surface (W/m^2) (A)
Large scale condensation heating rate at 28 model levels (K/s) (A)
Deep convective heating rate at 28 model levels (K/s) (A)
Deep convective moistening rate at 28 model levels (Kg/Kg/s)
Shallow convective heating rate at 28 model levels (K/s)
Shallow convective moistening rate at 28 model levels (Kg/Kg/s)
Vertical diffusion heating rate at 28 model levels (K/s)
Vertical diffusion zonal acceleration at 28 model levels (m/s^2)
Vertical diffusion meridional acceleration at 28 model levels (m/s^2)
Vertical diffusion moistening rate at 28 model levels (Kg/Kg/s)
Solar radiative heating rate at 28 model levels (K/s)
Long wave radiative heating rate at 28 model levels (K/s)

(A) .. Average in 6 hr forecast
(I) .. Instantaneous value at forecast hour 6
(Acc) .. Accumulation in 6 hour forecast
*) Not a predicted parameter but changes because of the time interpolation of the monthly climatology and snow cover change

Note that some parameters are redundant with surface flux file

4. Diagnostic file at FT=0 (on model Gaussian grid 192 x 94)

Surface skin temperature (K) (I)
Soil moisture content (kg/m^2) (I)
Water equivalent of accumulated snow depth (kg/m^2) (I)
Soil temperature at 10, 50 and 500 cm (K) (I)
Surface roughness (m) (I) *

(I) Instantaneous values
*) Not a predicted parameter but changes because of the time interpolation of monthly climatology

5. Grib sigma file (on model Gaussian grid 192 x 94)

Relative vorticity at 28 model levels (1/s)
Divergence at 28 model levels (1/s)
Thermodynamic temperature (not virtual) at 28 model levels (K)
Specific humidity at 28 model levels (kg/kg)
x-gradient of log surface pressure (1/m)
y-gradient of log surface pressure (1/m)
u wind (m/s) at 28 model levels
v wind (m/s) at 28 model levels
Surface pressure (Pa)
Surface geopotential height (gpm)
x-gradient of height (m/m)
y-gradient of height (m/m)

6. Restart file (no grib packing is done to insure reproducibility).

6.1 Sigma file (T62 L28) in spectral coefficient form

Vorticity at 28 model levels (1/s)
Divergence at 28 model levels (1/s)
Virtual temperature at 28 model levels (K)
Specific humidity at 28 model levels (Kgm/Kgm)
Log of surface pressure
Surface geopotential

6.2 Surface file (Gaussian grid 192 x 94)

Surface temperature
Soil moisture
Snow depth water equivalent
10 cm deep soil temperature
50 cm deep soil temperature
500 cm deep soil temperature
Surface roughness
Convective cloud cover

Convective cloud bottom height
Convective cloud top height
Land/sea/ice mask
Plant evaporation resistance
Ratio between 10 m winds and lowest model level winds

6.3 SST analysis and snow-ice analysis (Regular lat/lon grid)

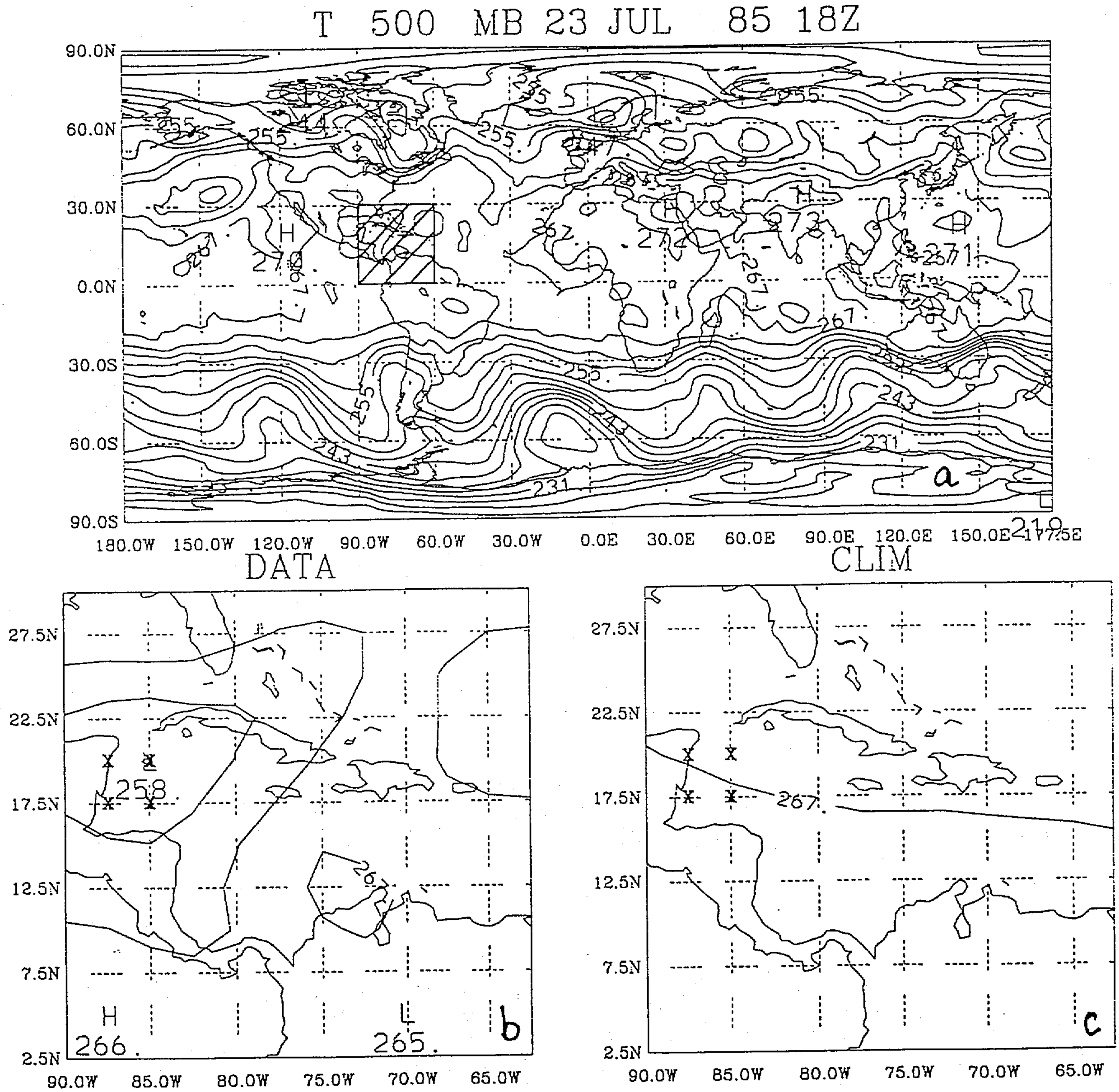
Sea surface temperature analysis 1 degree by 1 degree latitude-longitude grid
Snow-ice analysis 2 degree by 2 degree latitude-longitude grid

(or 18 levels.) Mand. Press. levels

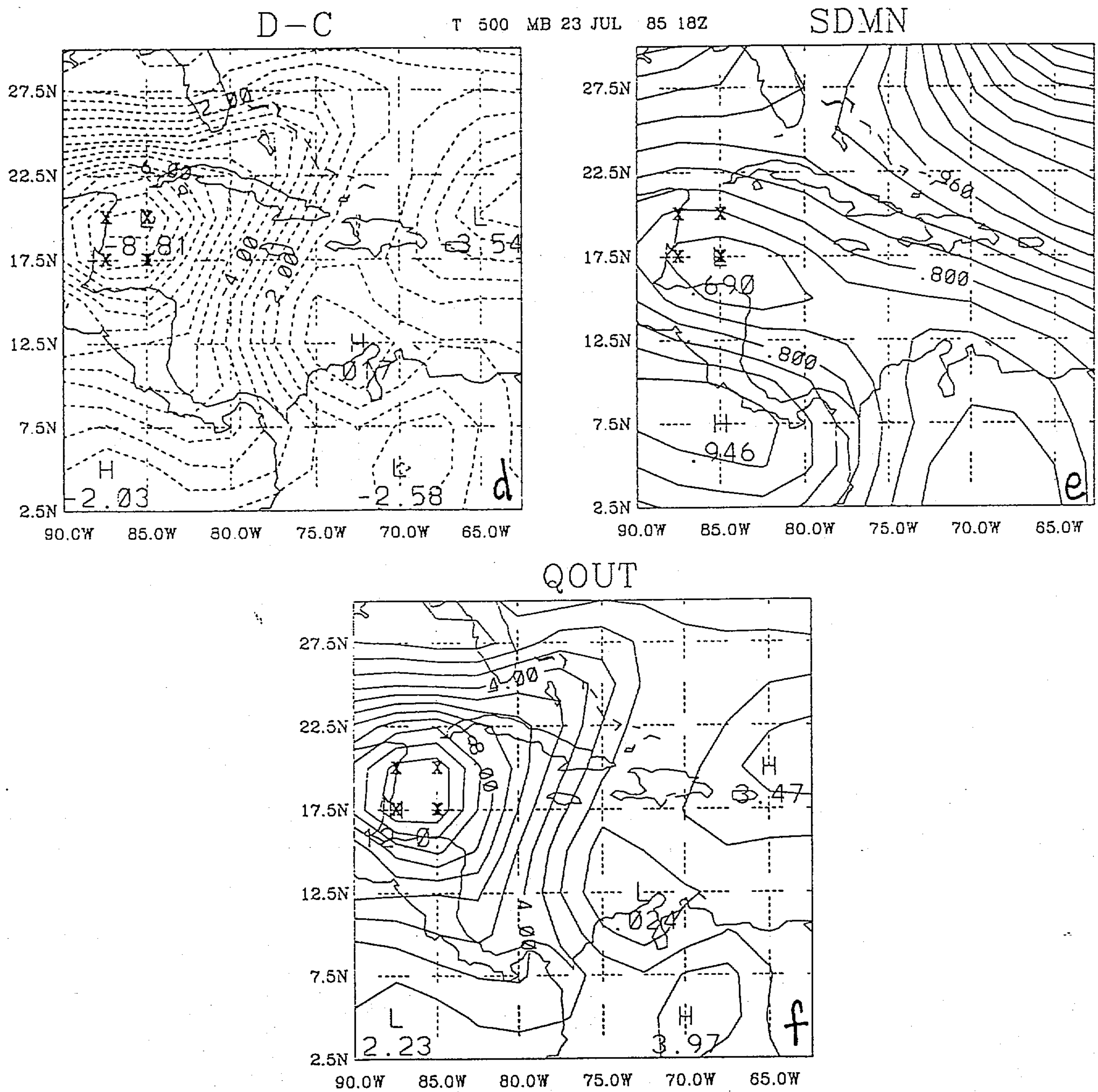
LEVEL	SIGMA	DELSIG	THICK (M)	OPSLEV	MANLEV
28	2.73	6.57			3.0
27	10.06	7.29	5599.		10.0
26	18.34	9.23	3828.	18	20.0
25	28.75	11.60	3053.		30.0
24	41.79	14.51	2621.		
23	58.05	18.03	2342.		50.0
22	78.15	22.22	2142.	17	70.0
21	102.78	27.09	1984.		100.0
20	132.61	32.62	1851.	16	
19	168.23	38.67	1729.	15	150.0
18	210.06	45.03	1612.	14	200.0
17	258.23	51.35	1495.	13	250.0
16	312.48	57.16	1376.	12	300.0
15	372.05	61.97	1260.	11	400.0
14	435.68	65.26	1139.	10	
13	501.68	66.69	1017.	9	500.0
12	568.09	66.06	895.	8	
11	632.90	63.47	776.		
10	694.26	59.19	664.	7	700.0
9	750.76	53.72	560.	6	
8	801.42	47.54	466.		
7	845.79	41.15	384.	5	850.0
6	883.84	34.93	313.		
5	915.92	29.19	253.	4	925.0
4	942.55	24.05	203.		
3	964.37	19.59	162.	3	
2	982.08	15.82	129.	2	
1	995.00	10.00	80.	1	

Table 1

VIII.a



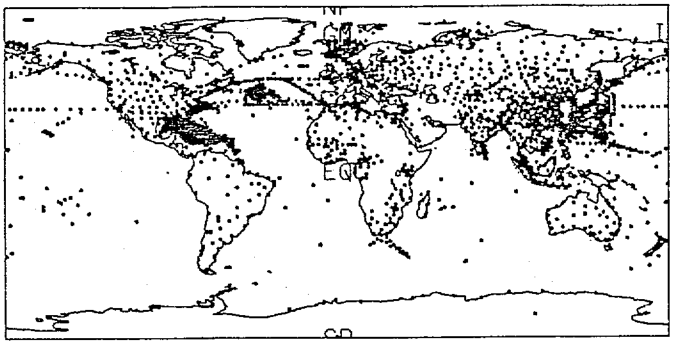
CONTOUR FROM 218.80 TO 272.80 CONTOUR INTERVAL OF 3.0000 PT(3,3)= 268.80



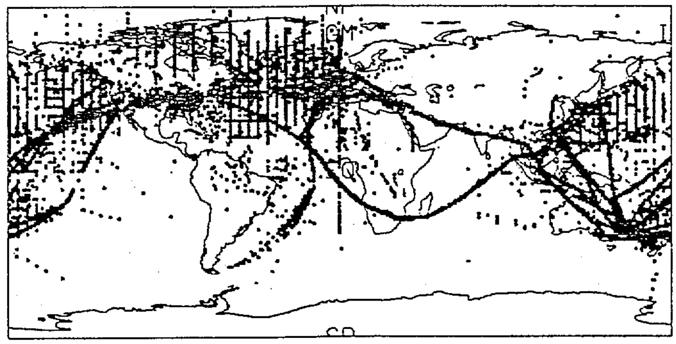
CONTOUR FROM 0800000E+00 TO 0100000E+00 CONTOUR INTERVAL OF 0150000E-01 PT(3,3)= 0395550

DATA DISTRIBUTION Ø1AUG85ØØZ-31AUG8518Z

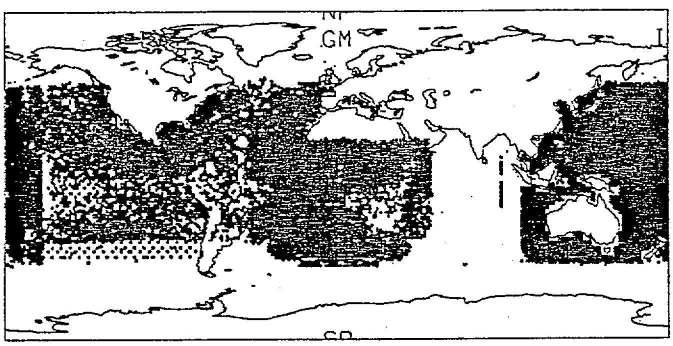
ADPUPA



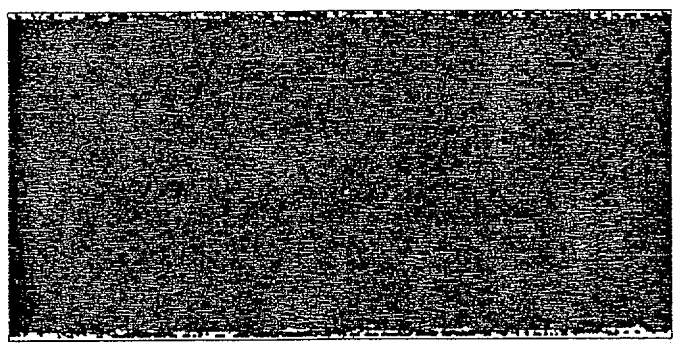
AIRCFT



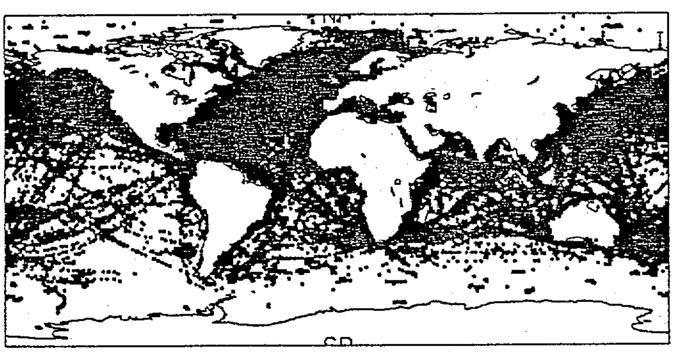
SATWND



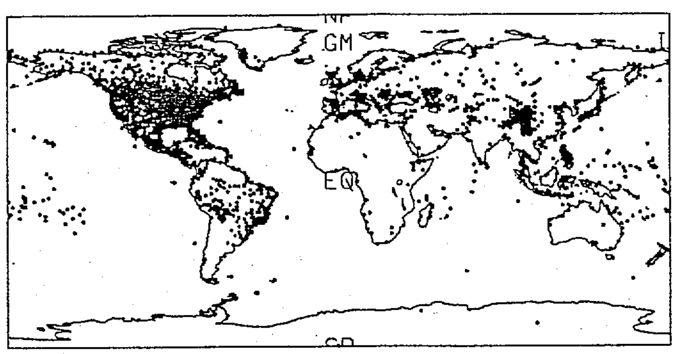
SATSND



ADPSHP

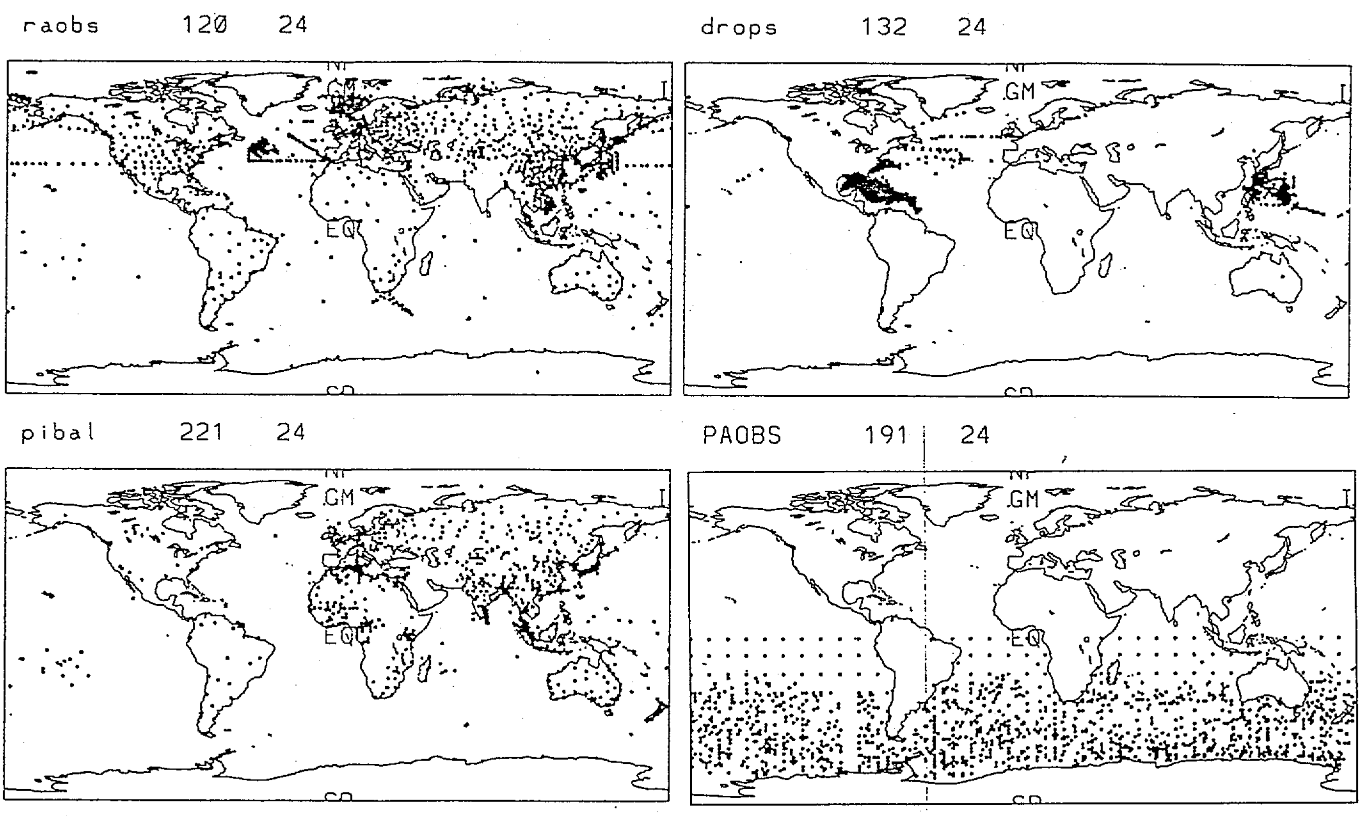


ADPSFC



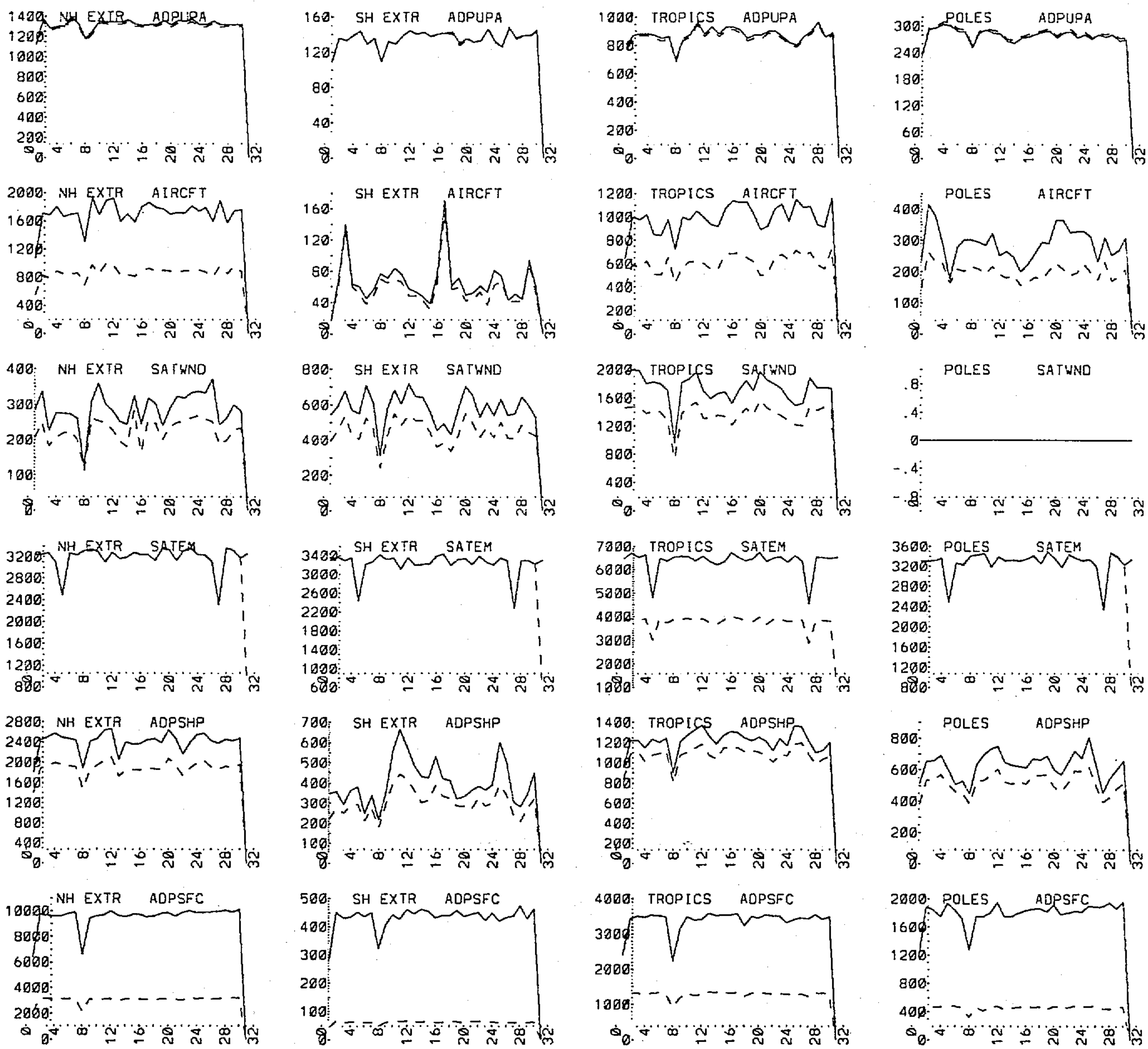
VIII 1a

DATA DISTRIBUTION Ø1AUG85- 31AUG85



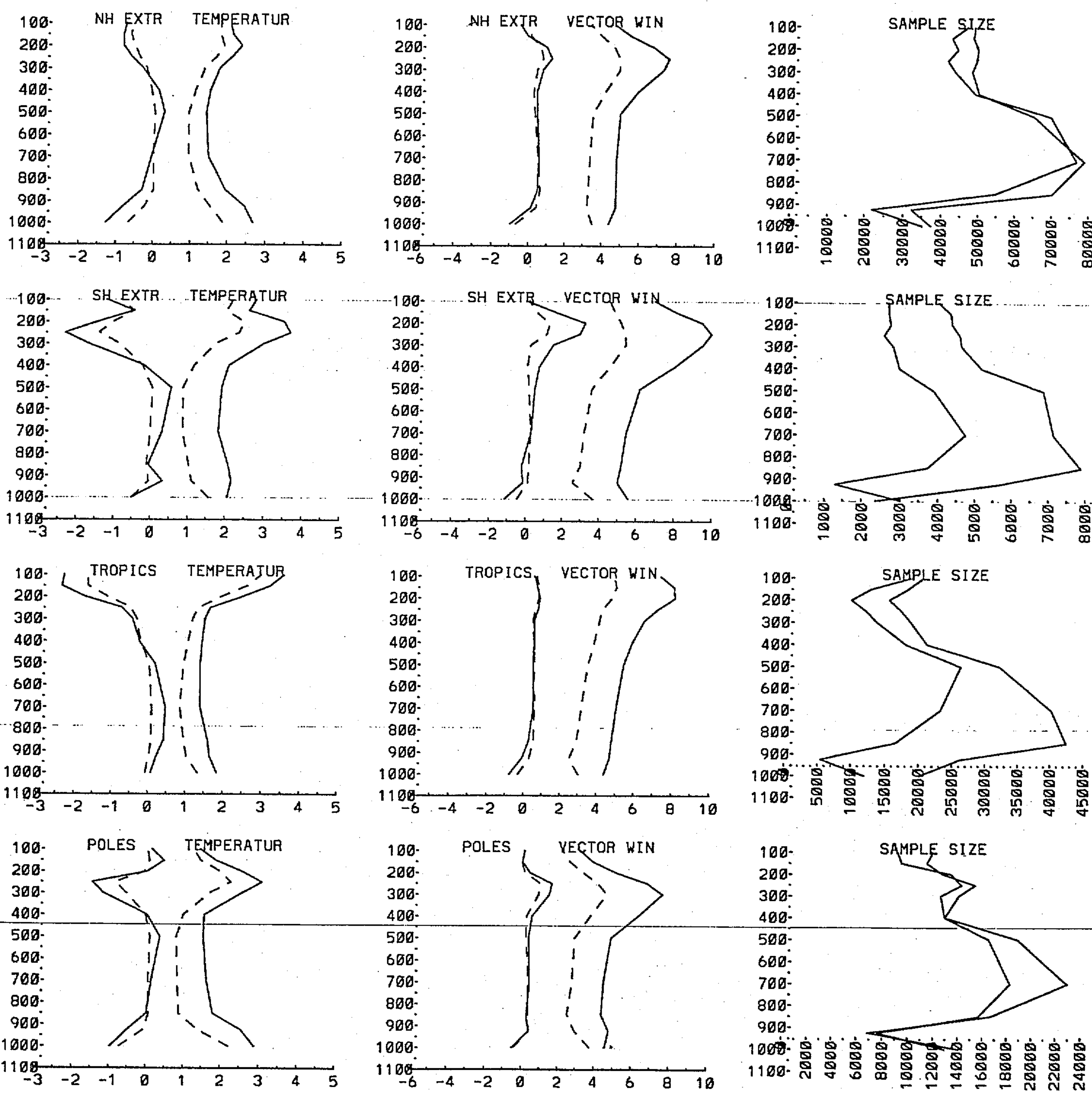
VIII 1b

DATA COUNTS FOR AUG85 ON29 vs PREPDA



pilot21

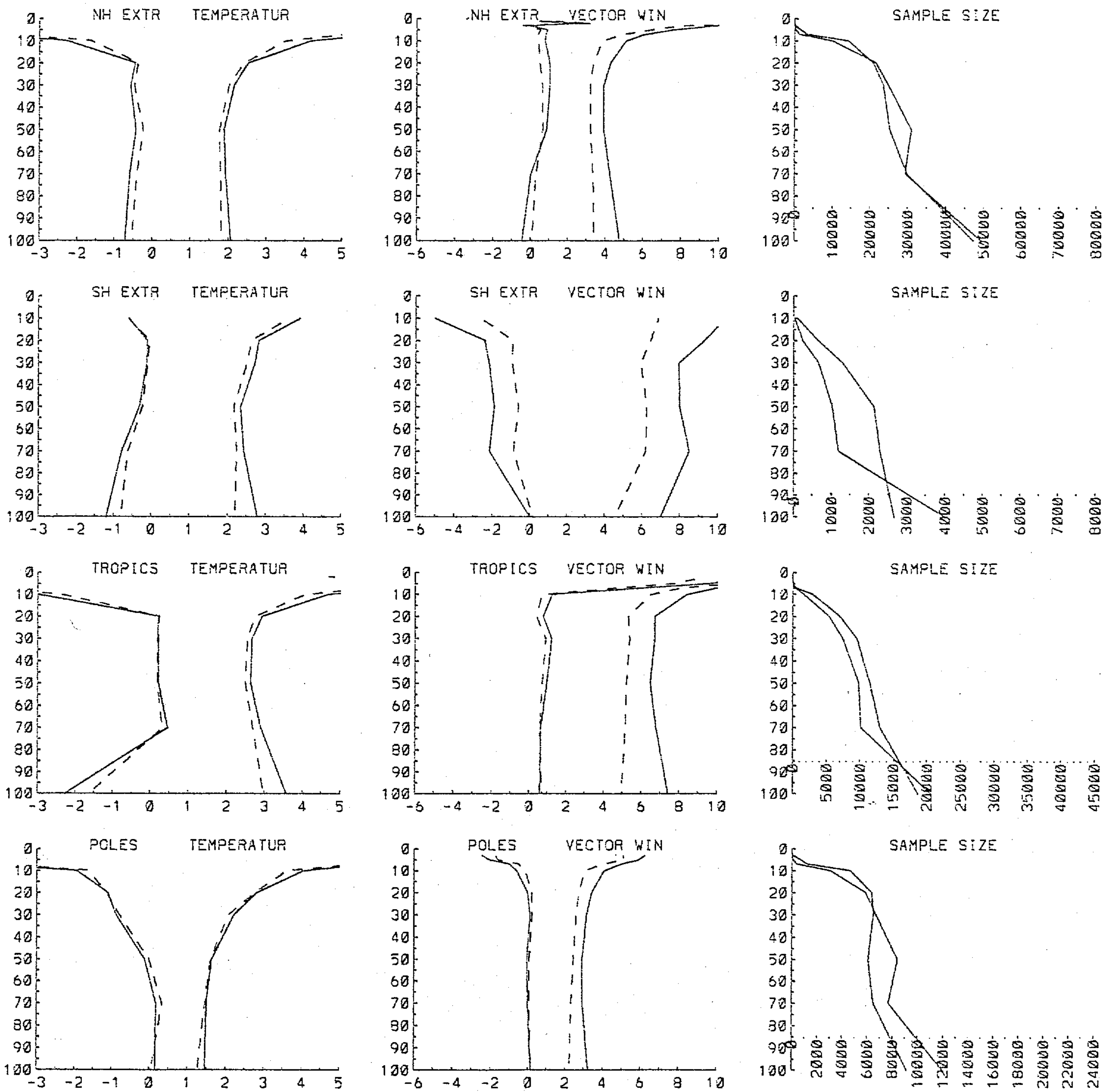
ADPUPA Ø1AUG85-31AUG85 SGES SANL



VIII 3a

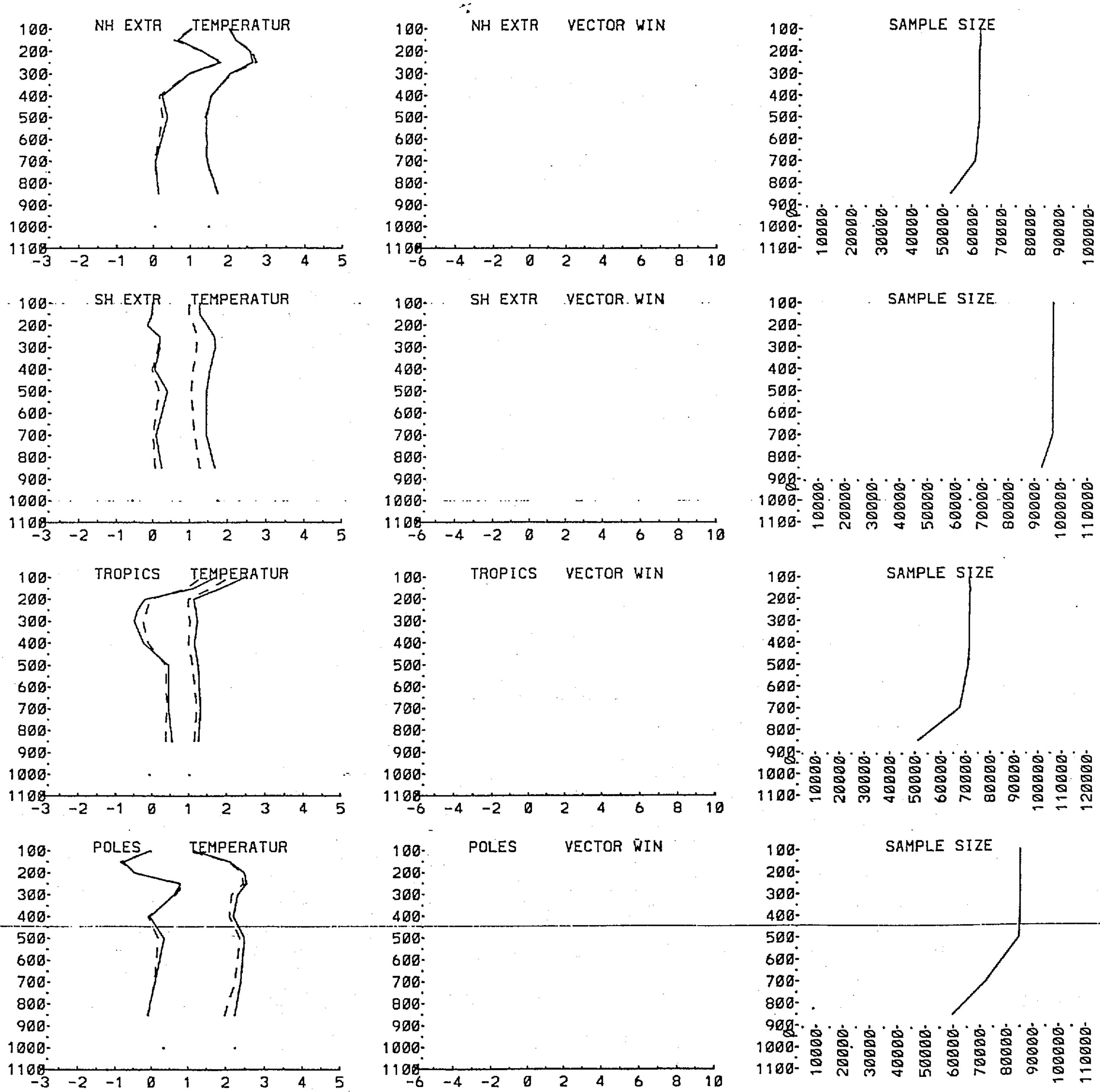
pilot21

ADPUPA 85080100-85083018 SGES SANL



pilot21

SATEMP Ø1AUG85-31AUG85 SGES SANL



VIII 3c

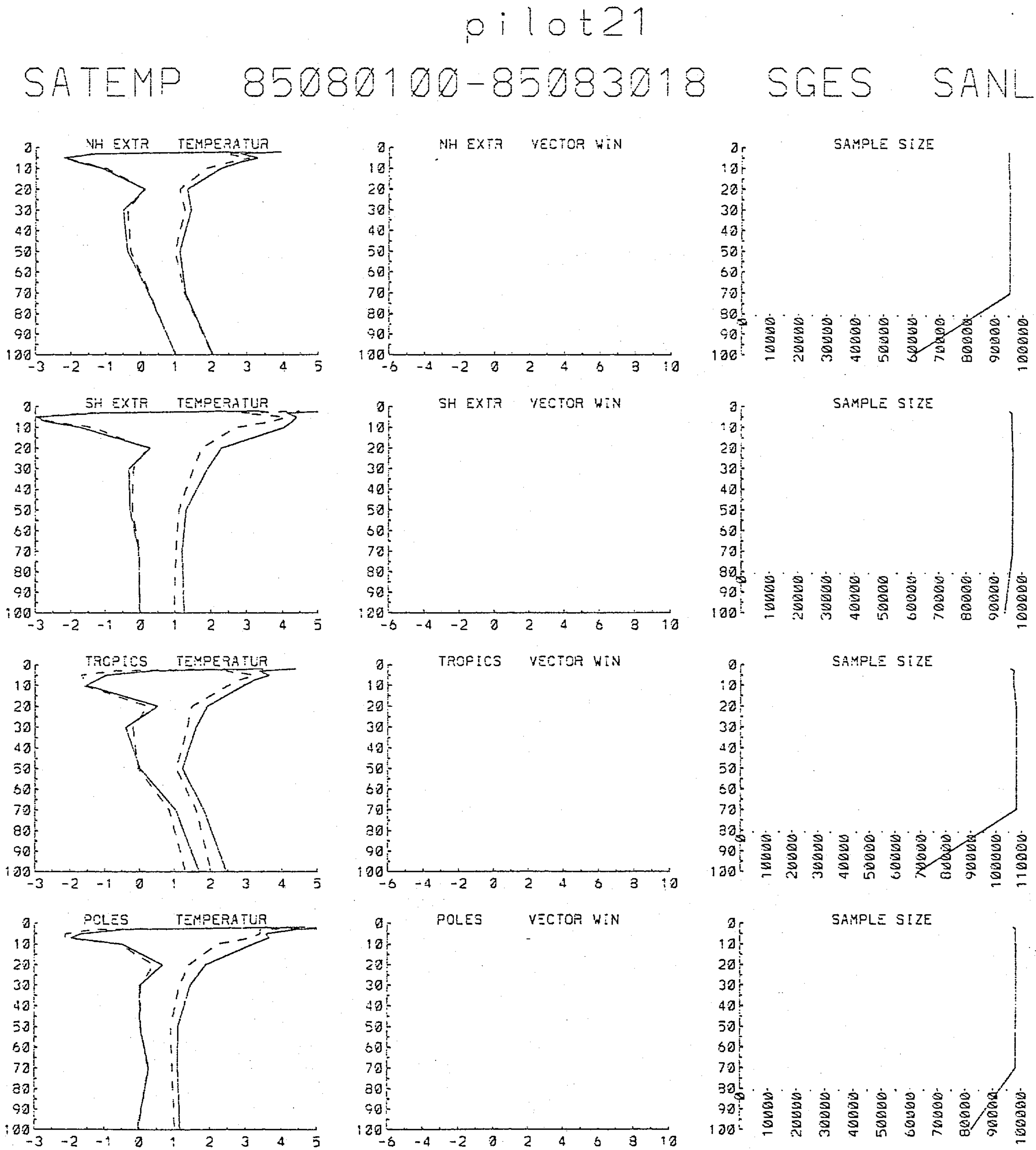
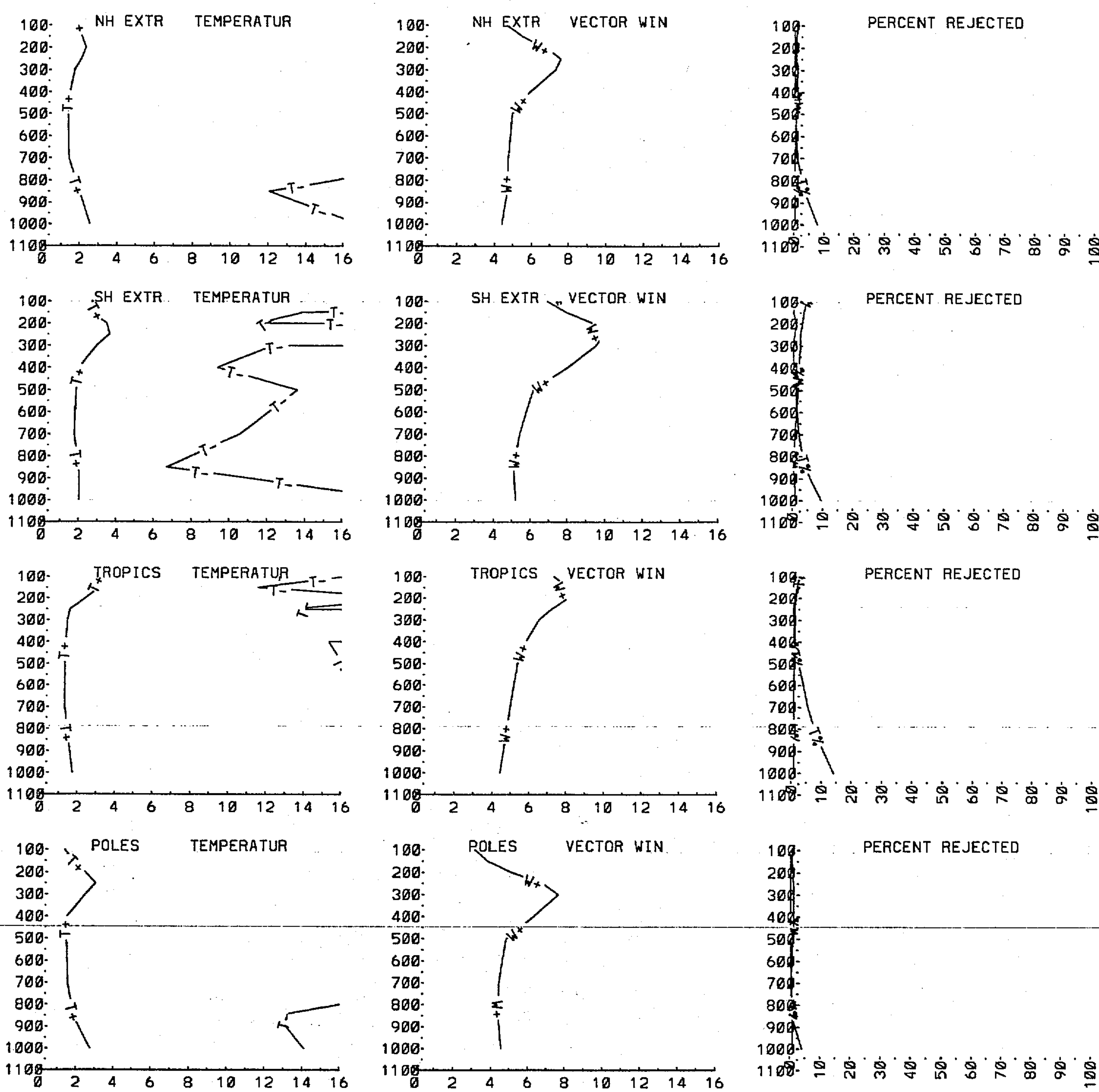


Figure 3d. Same as Figure 3c except 100-10 mb

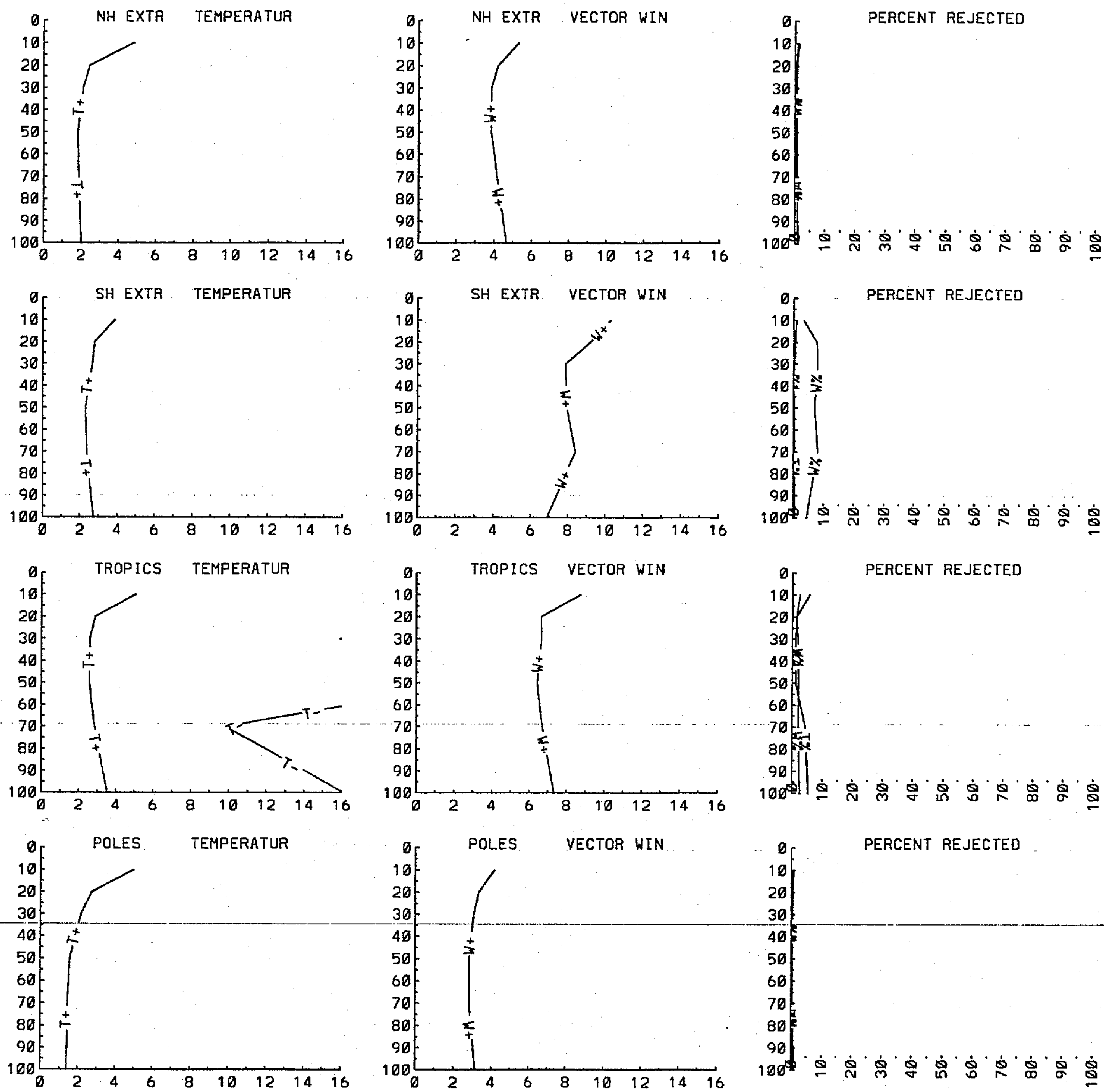
VIII 3d

ADPUPA 85Ø8Ø1ØØ-85Ø83118 PILOT21



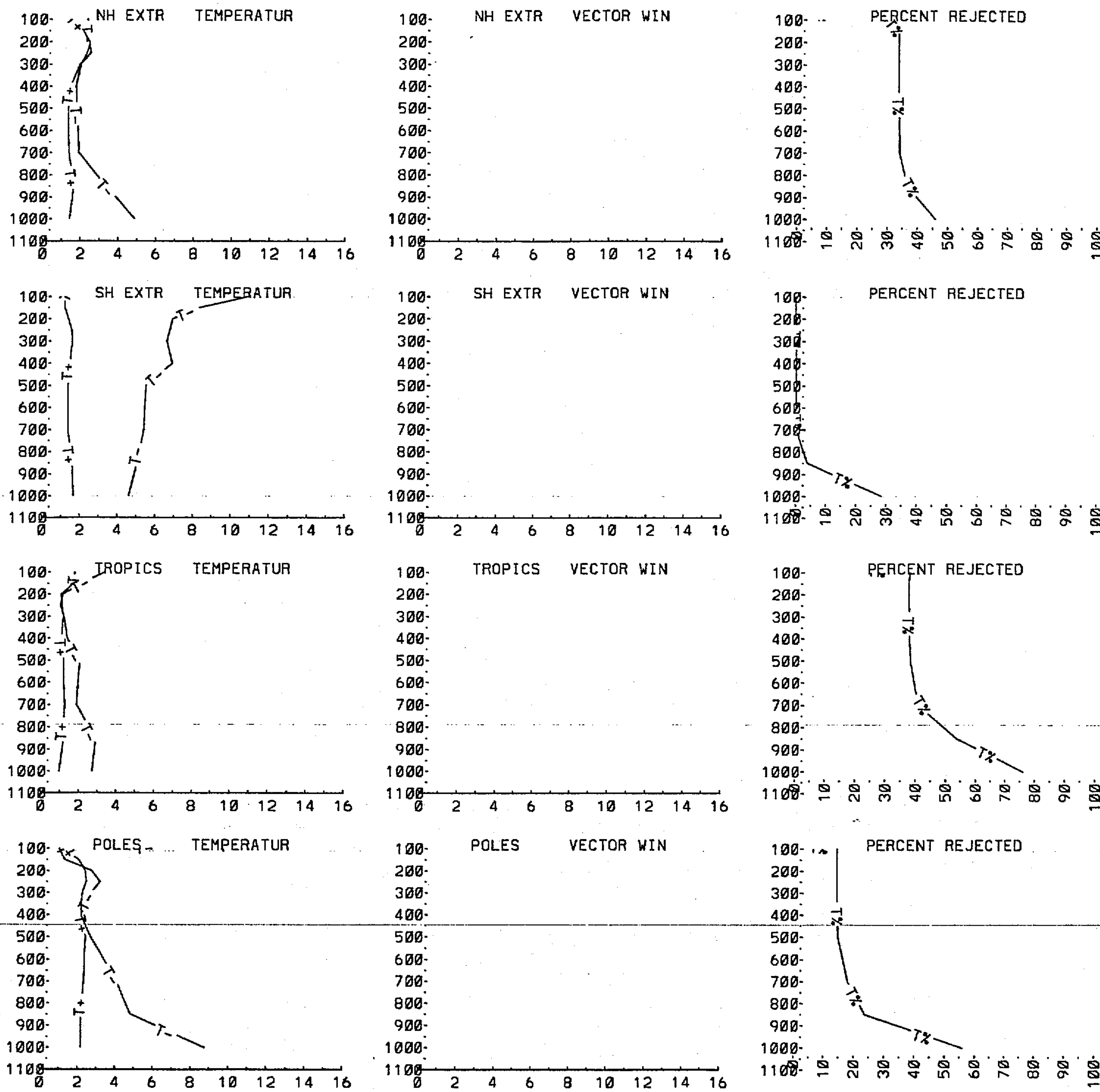
VIII 4a

ADPUPA 85Ø8Ø1ØØ-85Ø83118 PILOT21



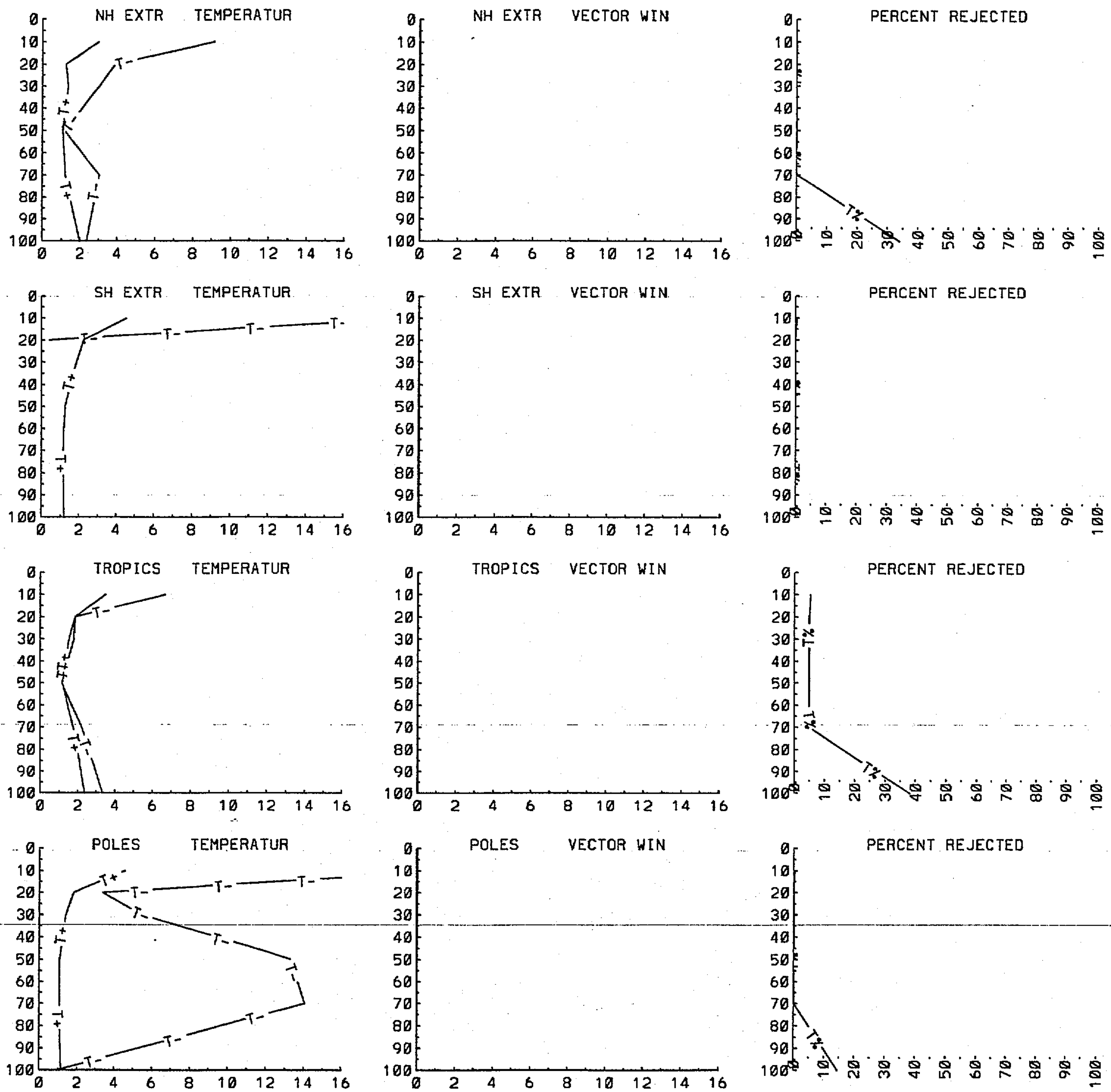
VIII 46

SATEMP 85Ø8Ø1ØØ-85Ø83118 PILOT21



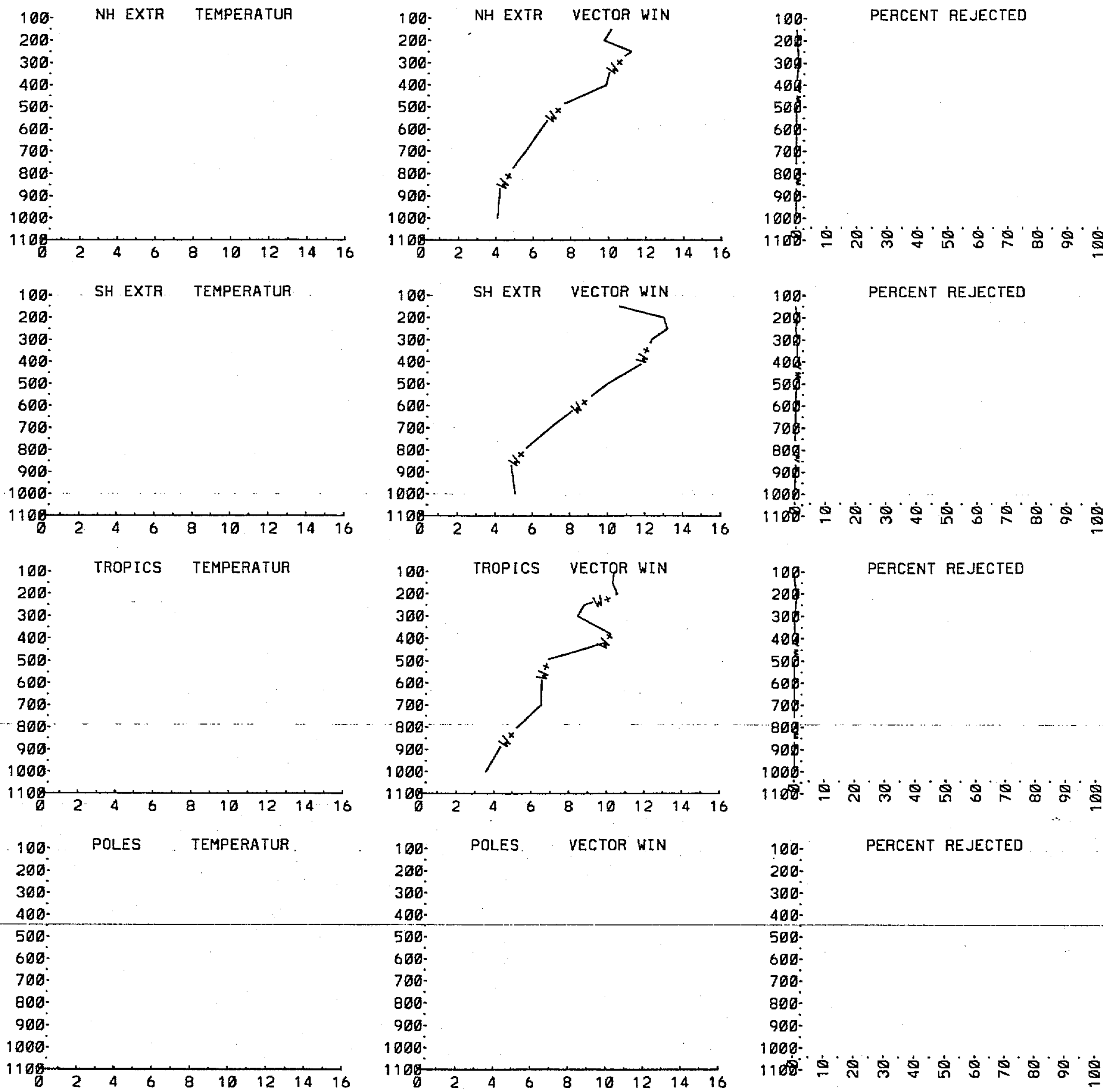
VIII 4c

SATEMP 85Ø8Ø1ØØ-85Ø83118 PILOT21



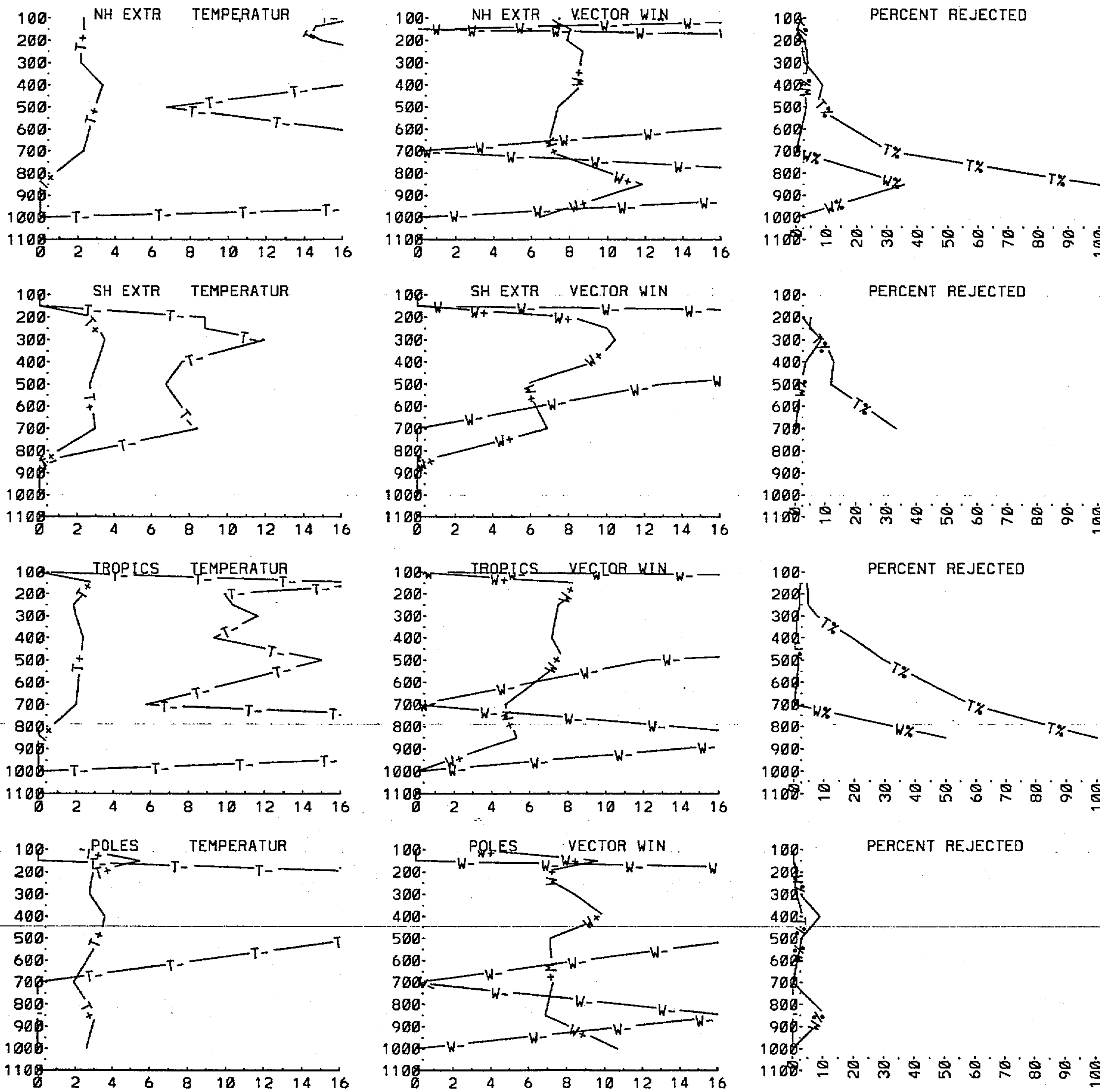
VIII 4d

SATWND 85Ø801ØØ-85Ø83118 PILOT21



VIII 4e

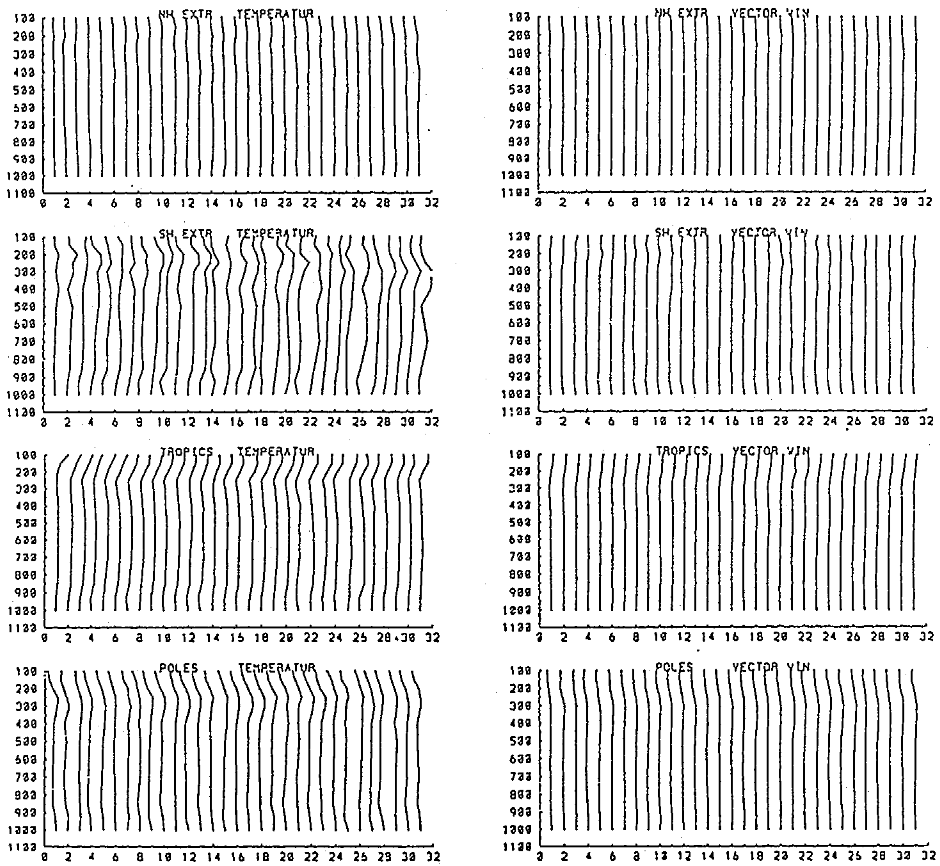
AIRCFT 85Ø8Ø1ØØ-85Ø83118 PILOT21



VIII 4f

pilot21

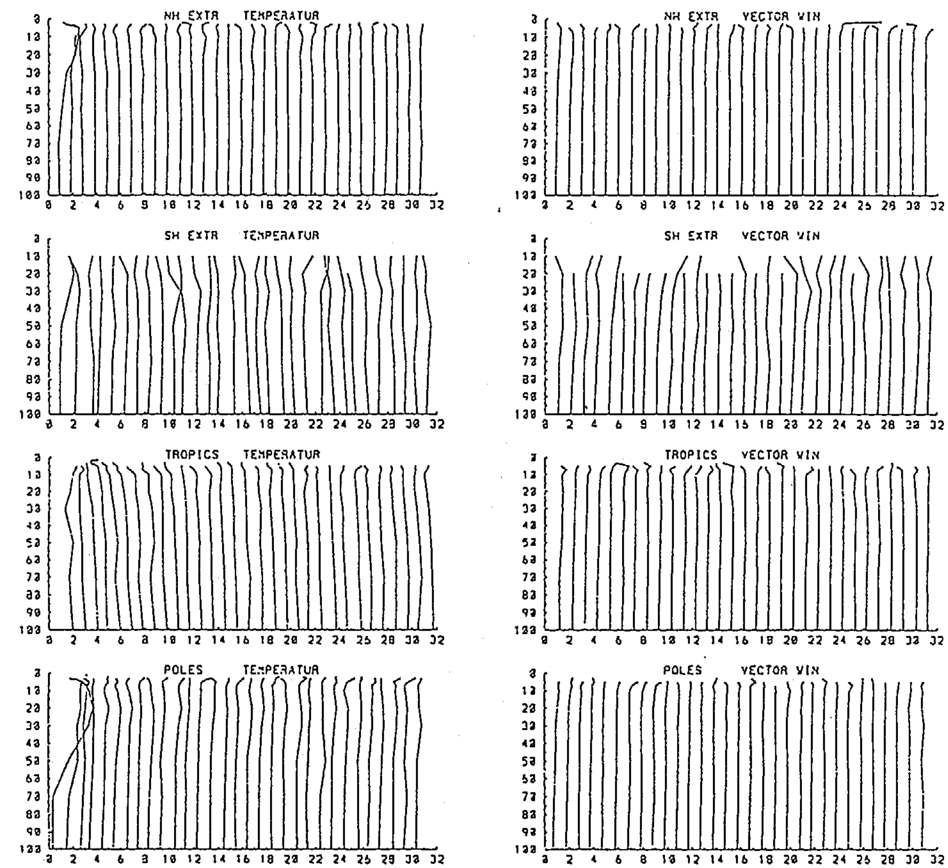
ADPUPA JUL85 PSGES



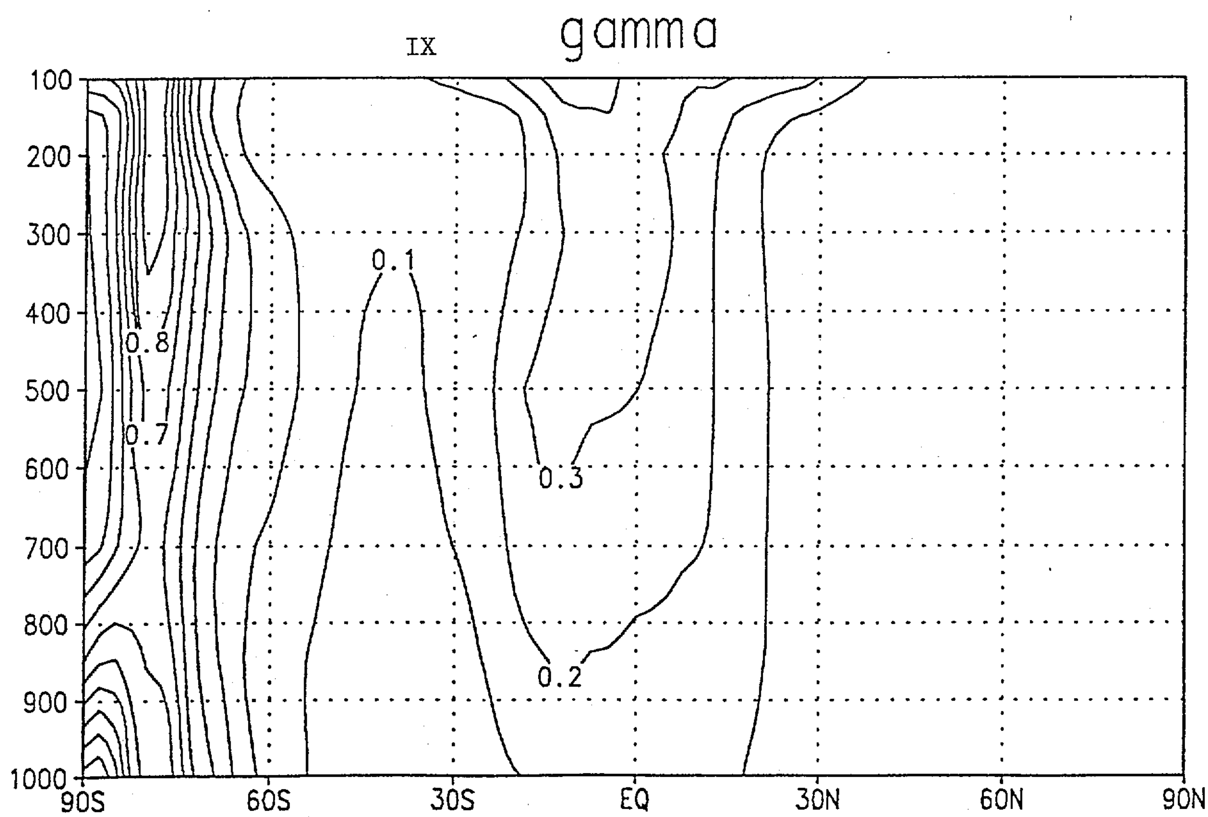
5a.

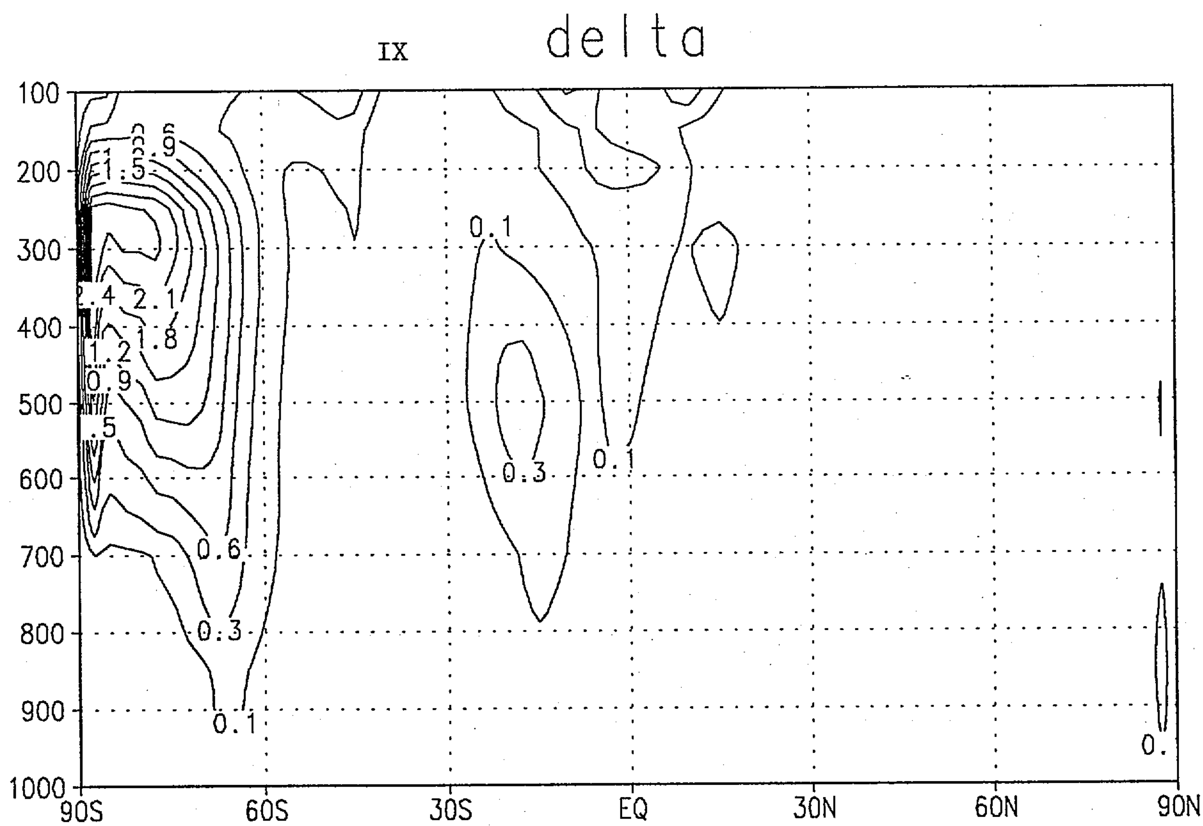
pilot21

ADPUPA JUL85 PSGES



5b.

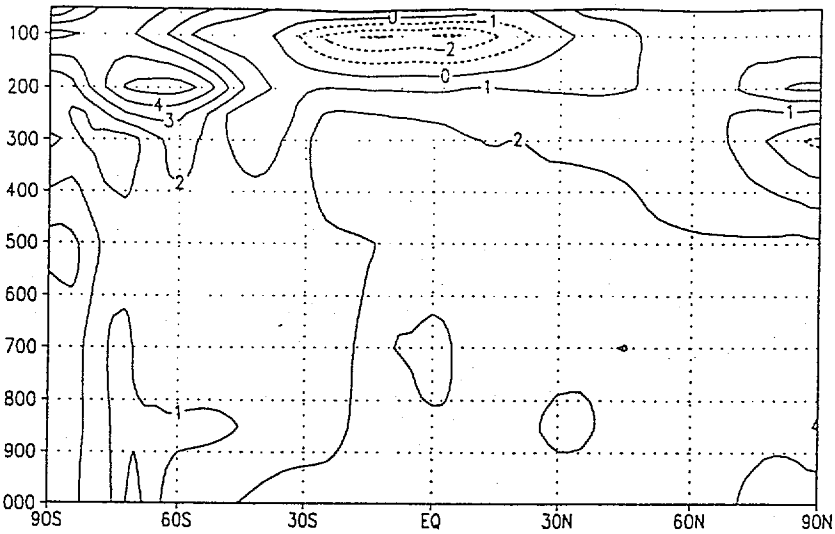




GrADS: COLA/UMCP

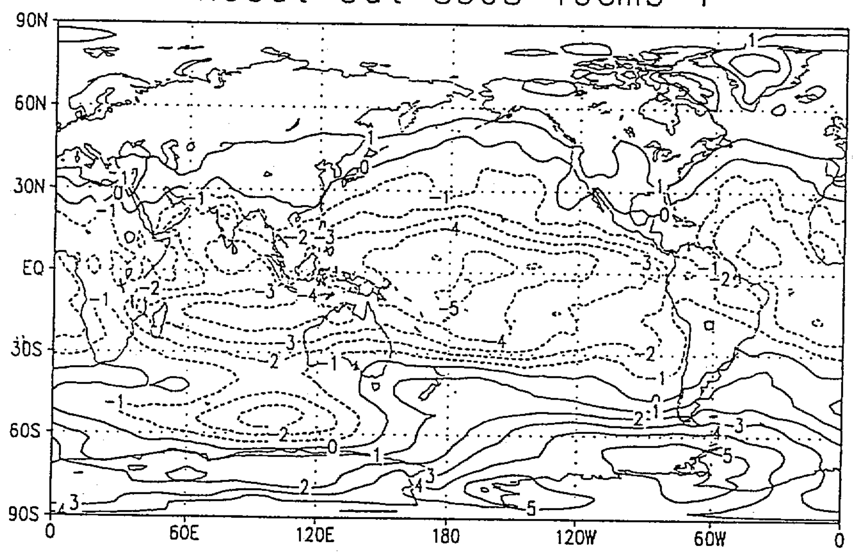
IX 1

nosat-sat 8508 T



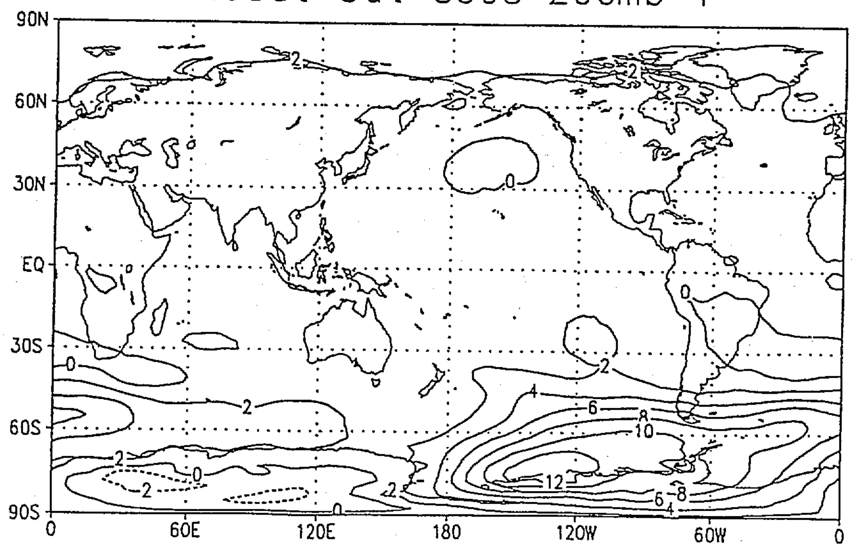
1a

nosat-sat 8508 100mb T



1b

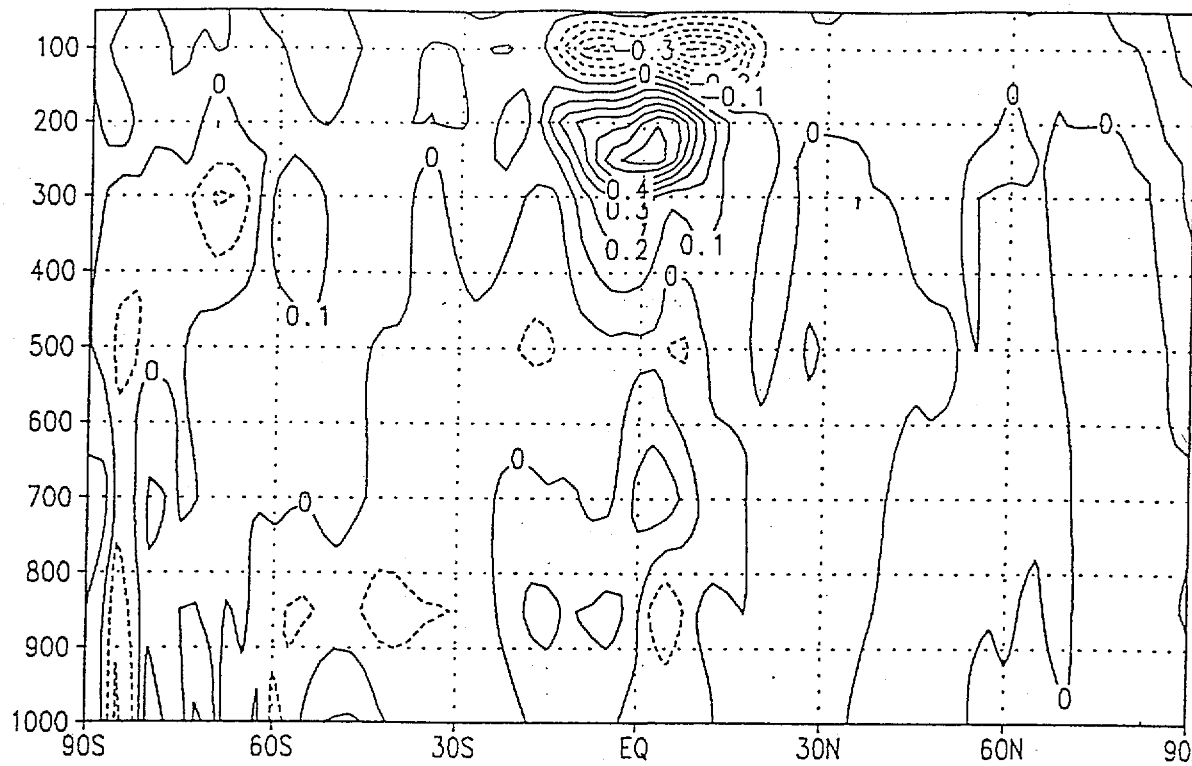
nosat-sat 8508 200mb T



1c

IX 1

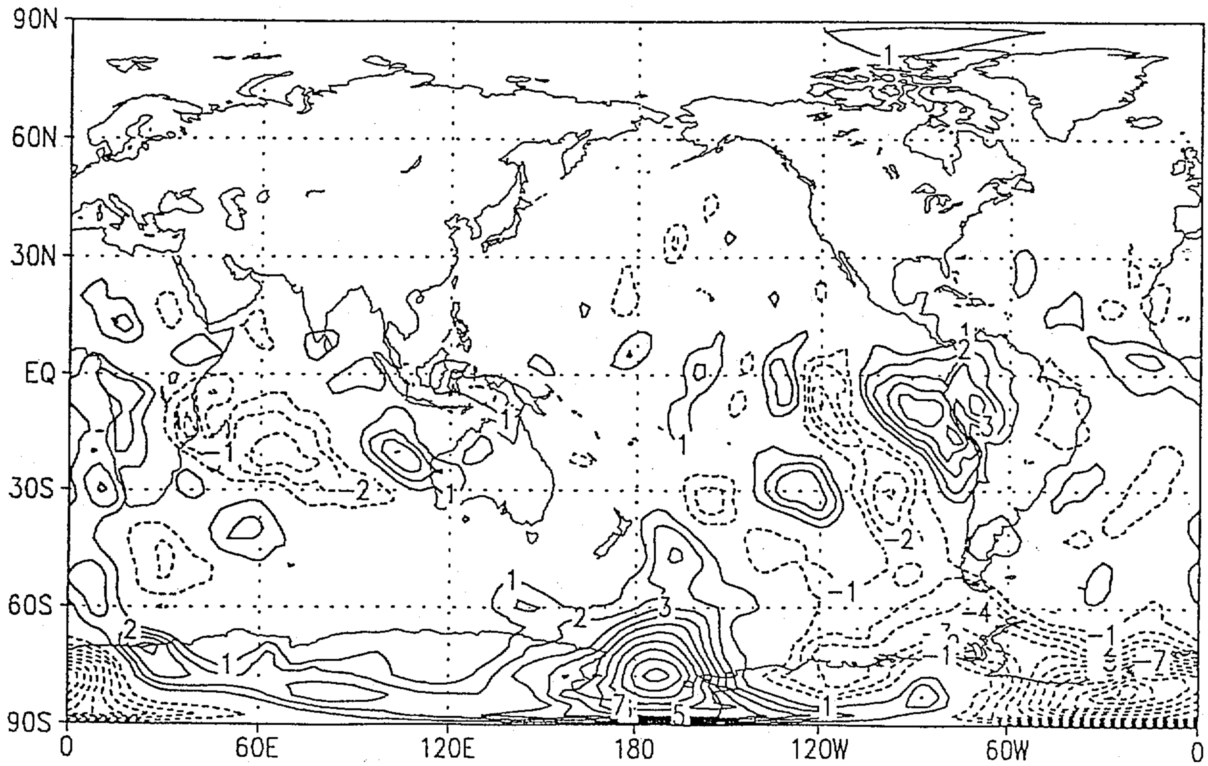
nosat-sat 8508 v



1d

2

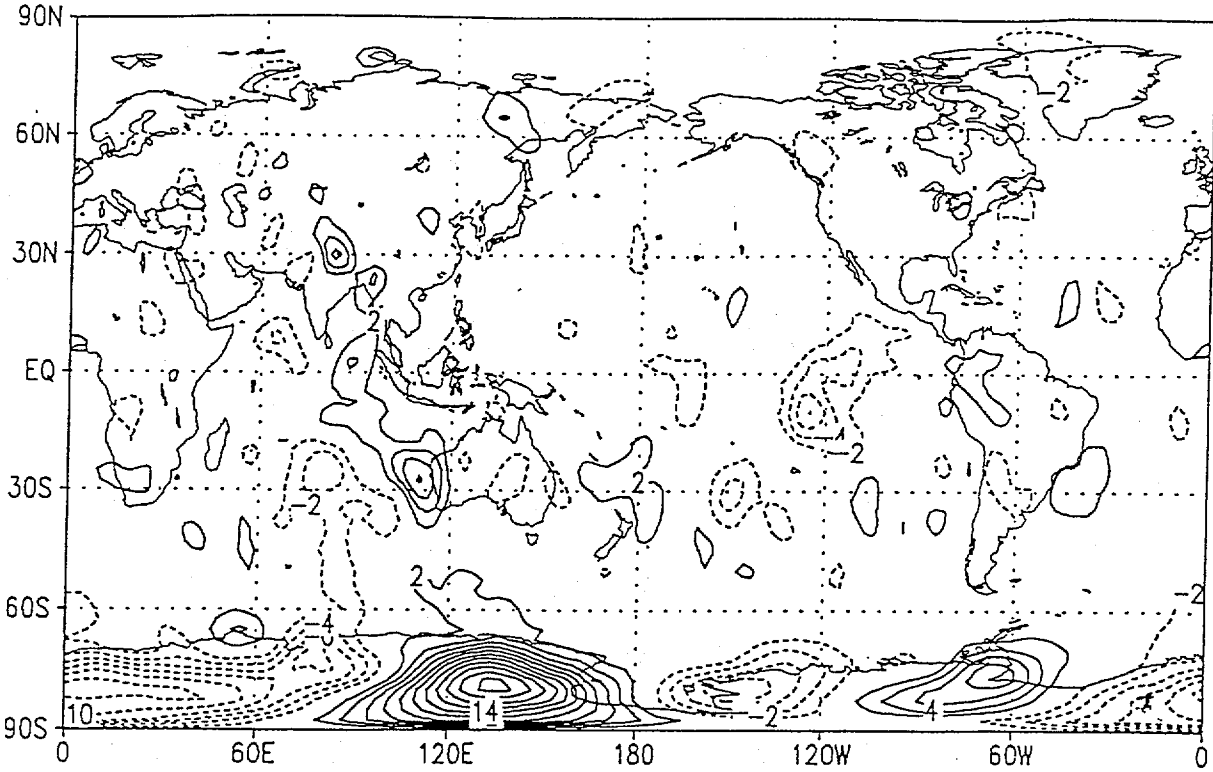
nosat-sat 8508 250mb v



1e

30

cddb-sat

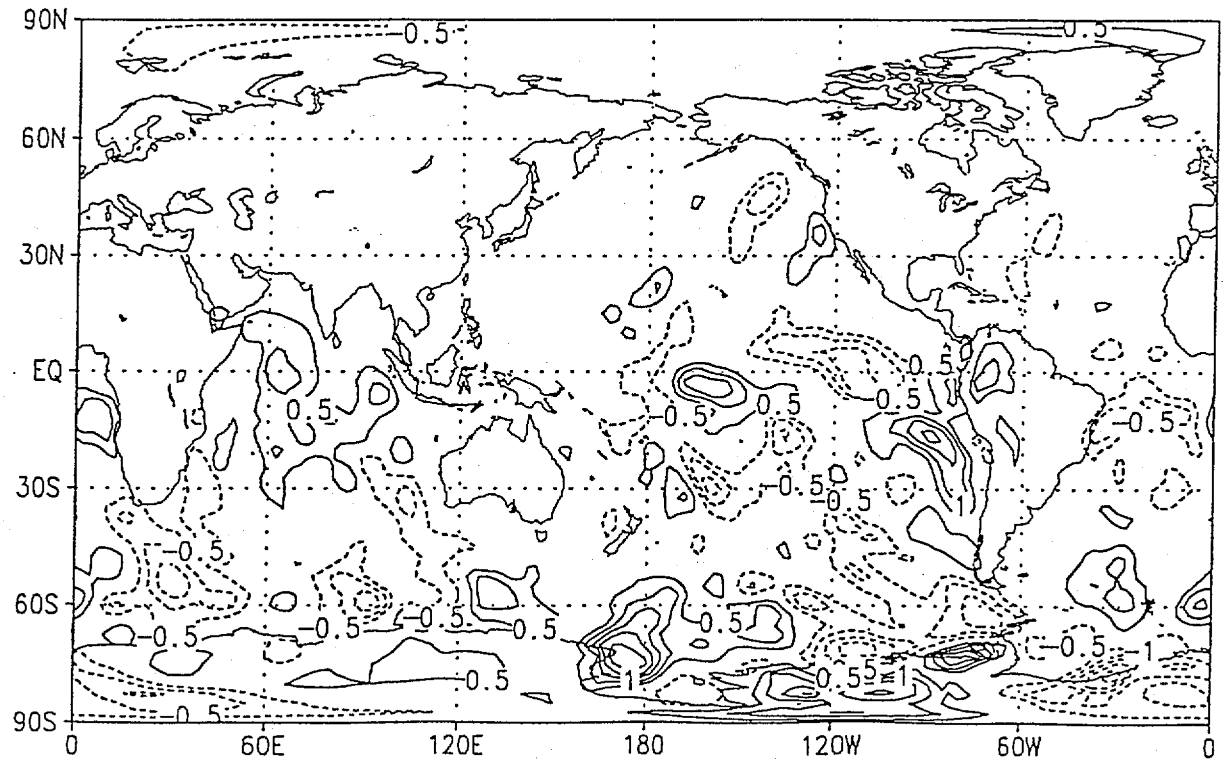


IX 1

1f

2e

nosat-sat 8508 850mb v

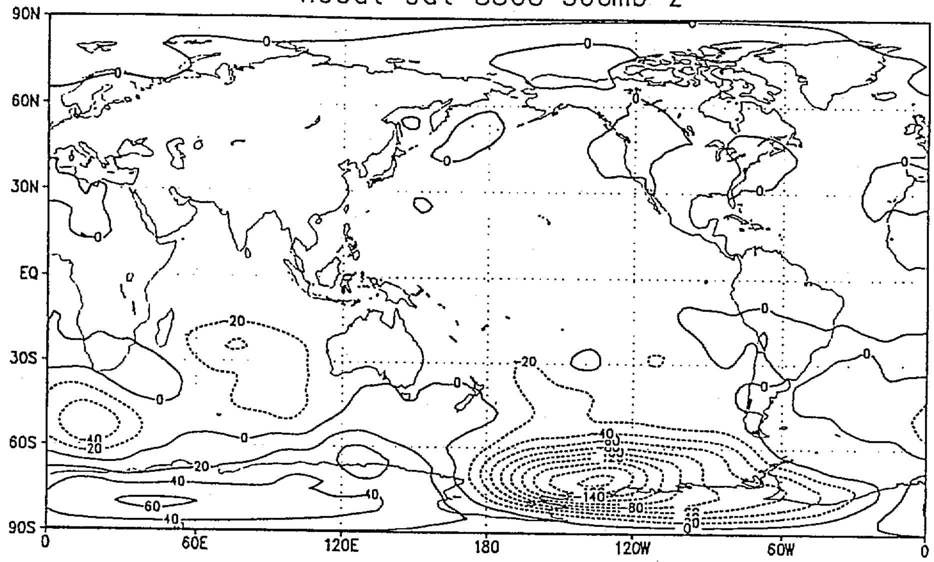


1g

2d

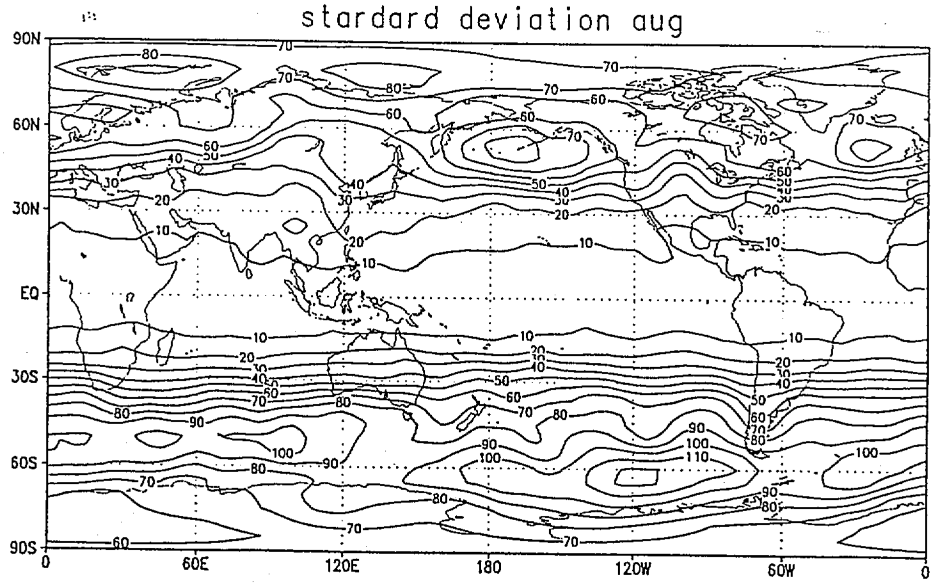
1X 2

nosat-sat 8508 500mb z



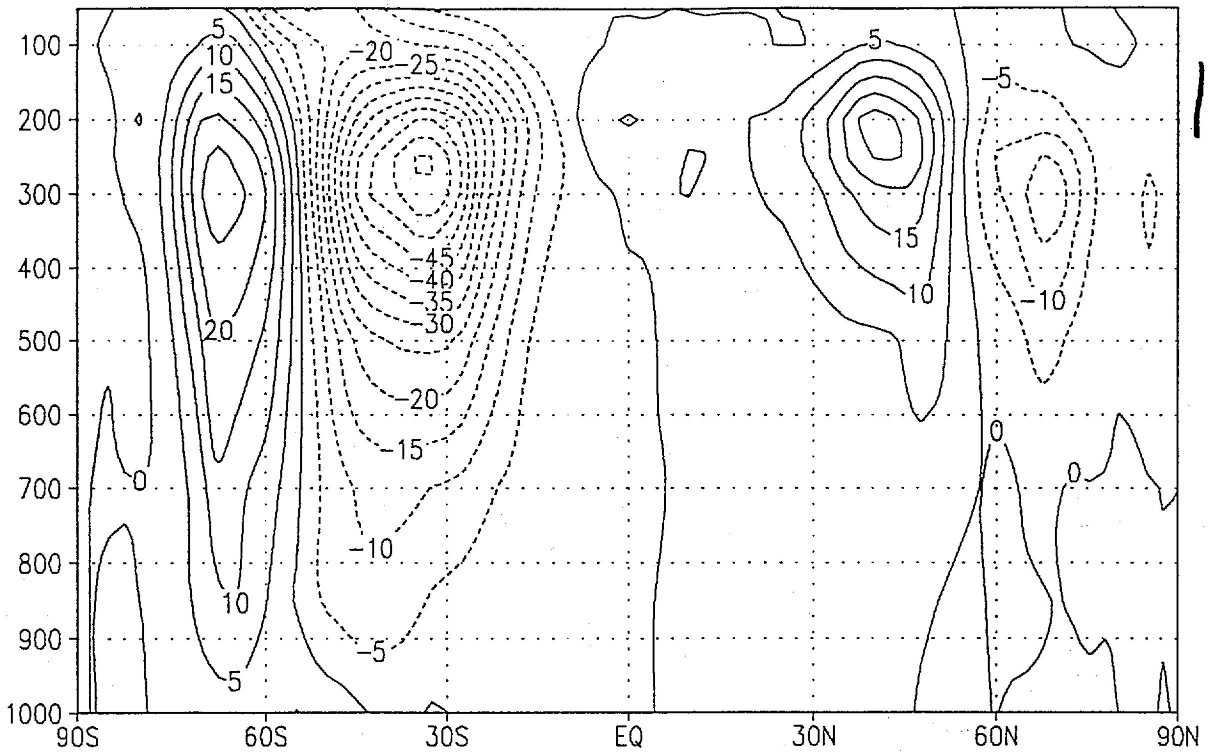
2a

stardard deviation aug



2b

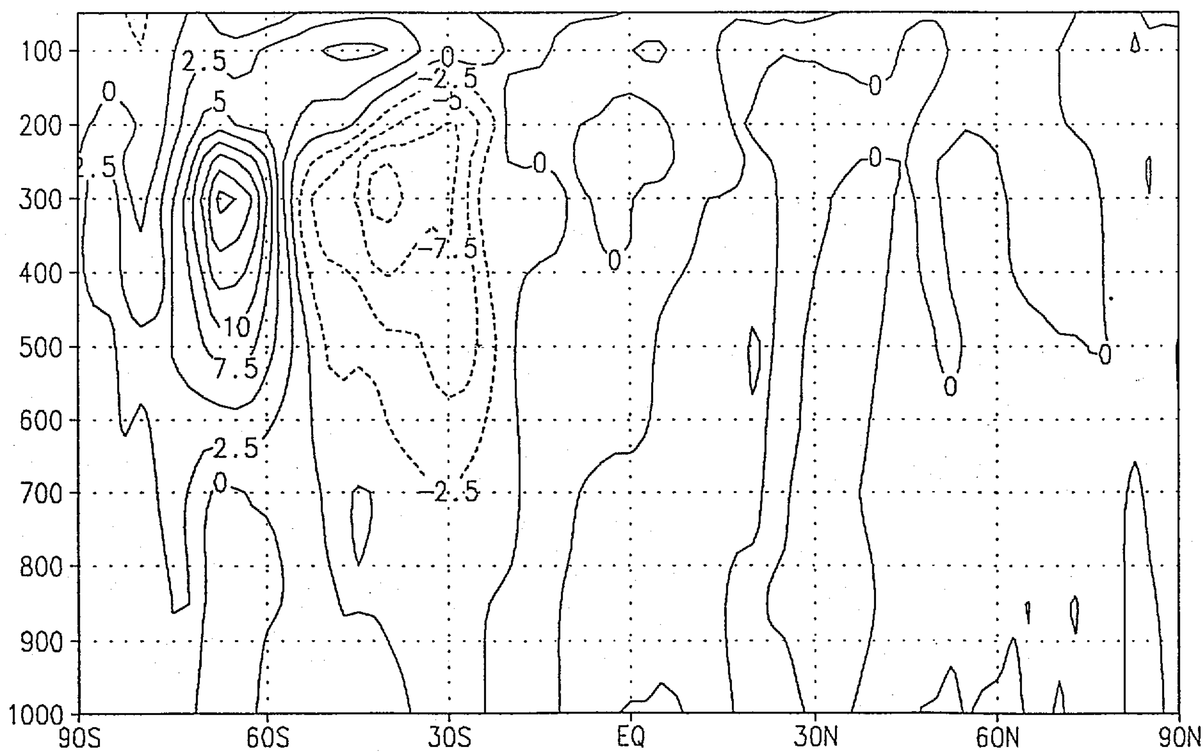
sat eddy $\overline{u'v'}$ 8508



1X2

2c

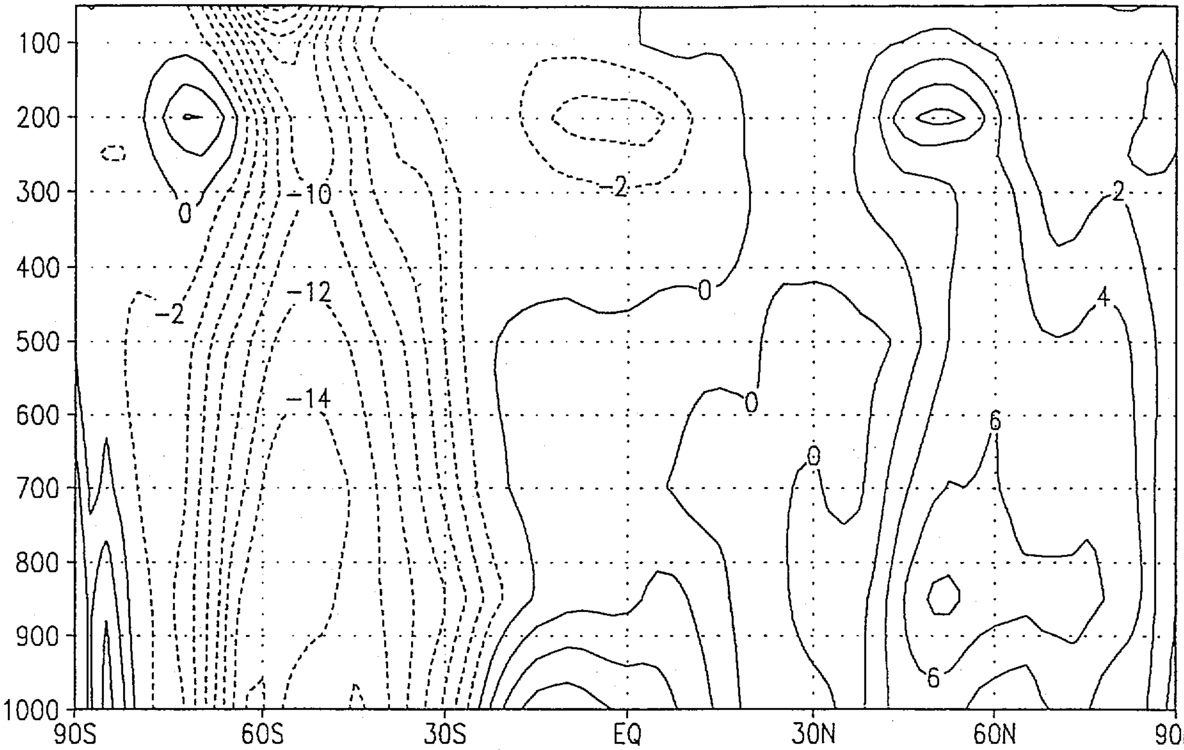
nosat-sat eddy $\overline{u'v'}$



2d

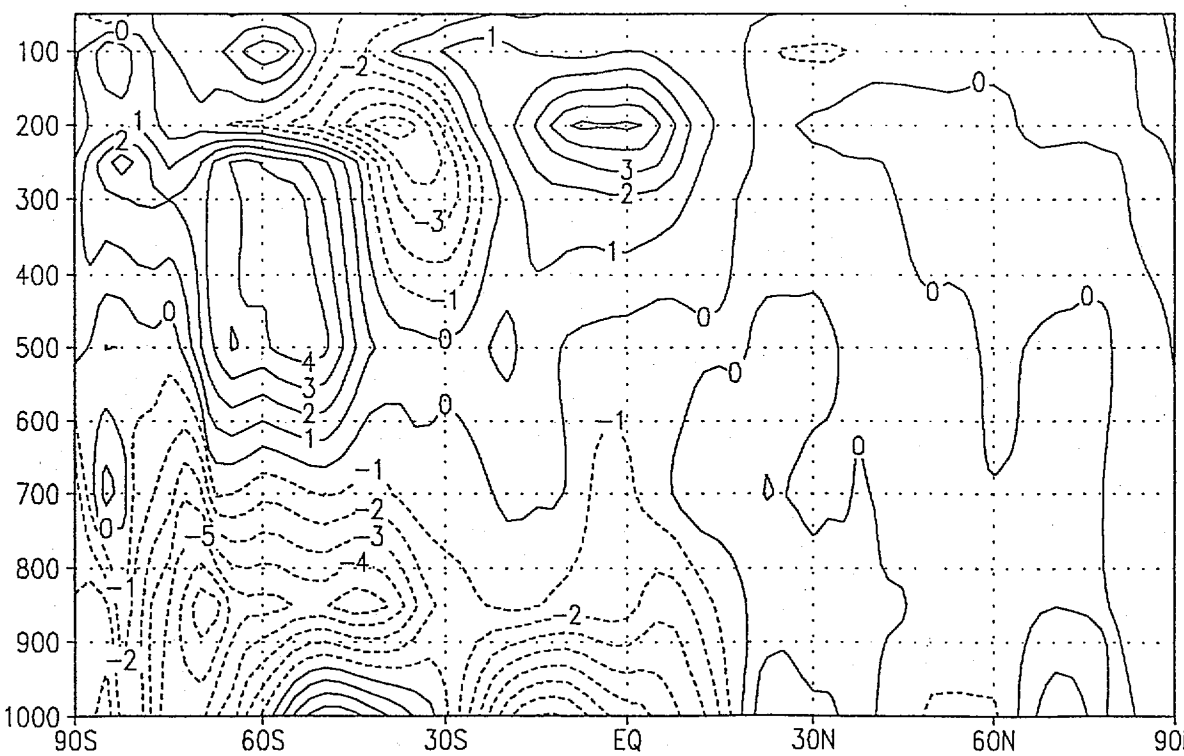
sat eddy $\overline{v't'}$ 8508

1X2



2e

nosat-sat eddy $\overline{v't'}$

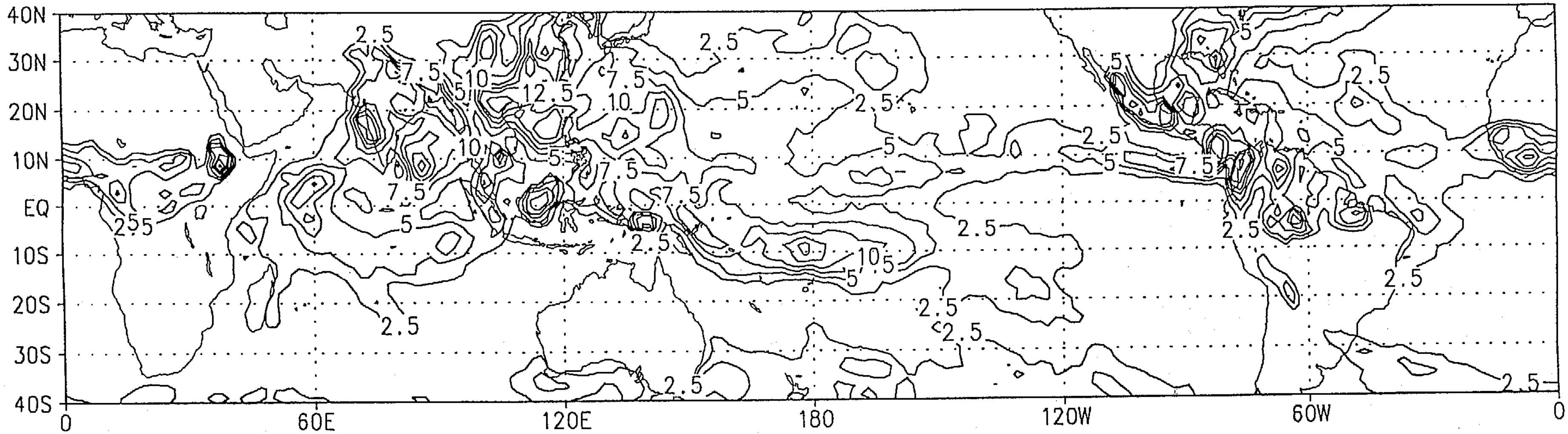


2f

1X3

3

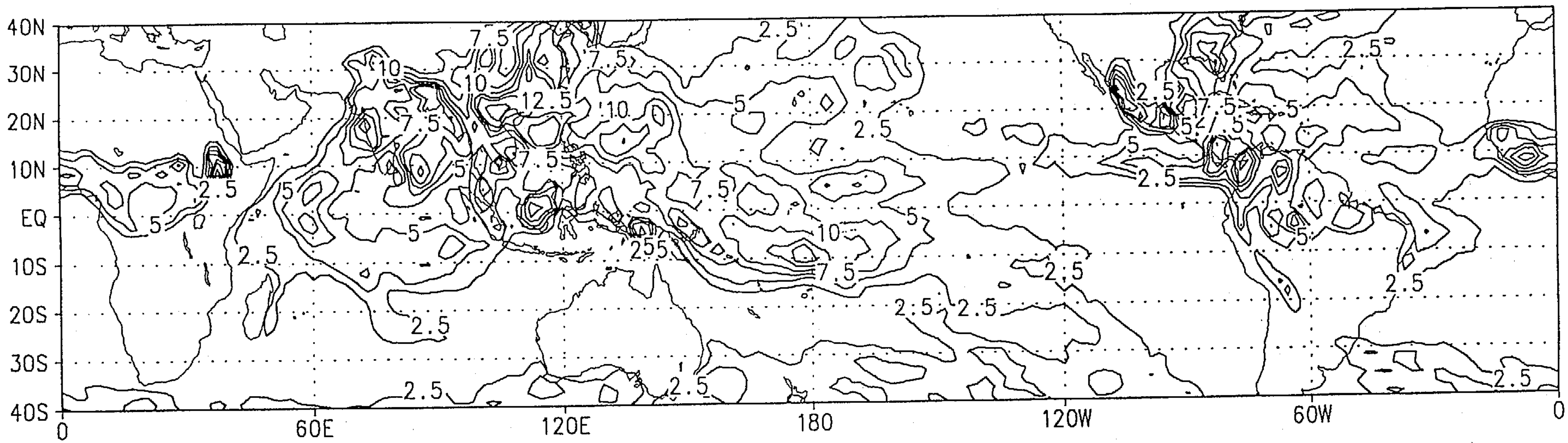
rain sat 8508



3a

GrADS: COLA/UMCP

rain nosat 8508

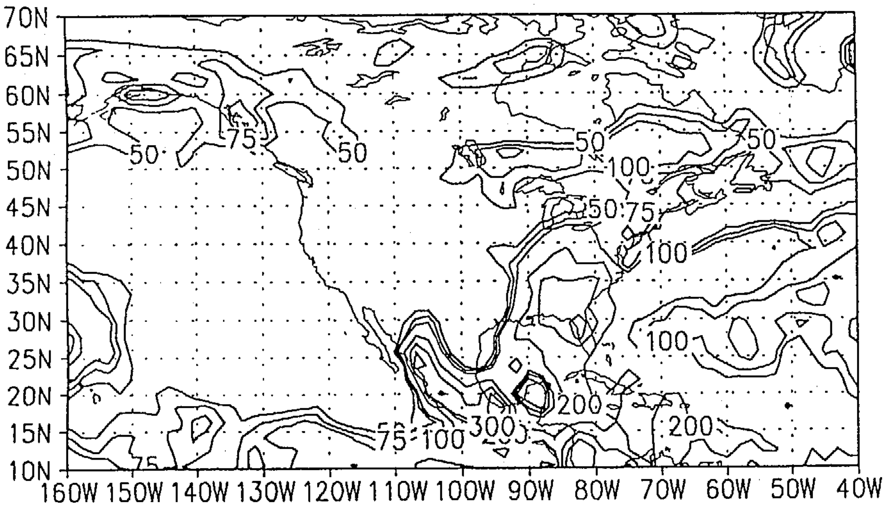


3b

GrADS: COLA/UMCP

IX 3

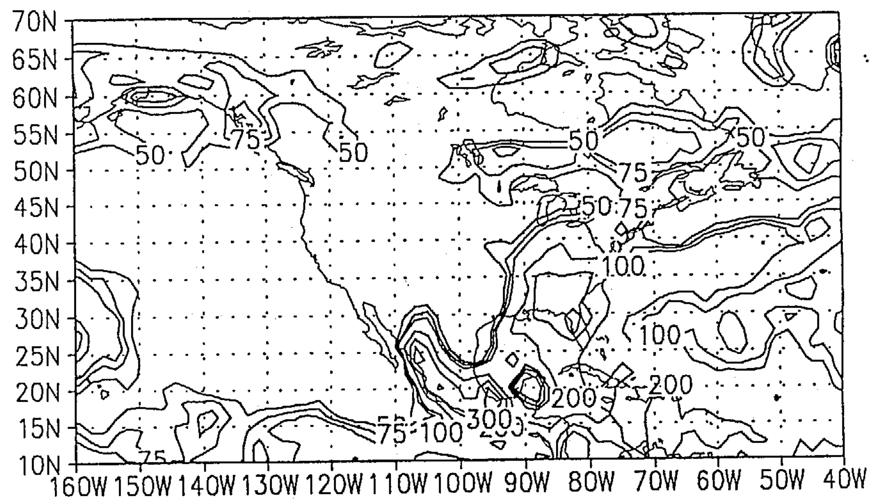
rain sat 8508



3c

GrADS: COLA/UMCP

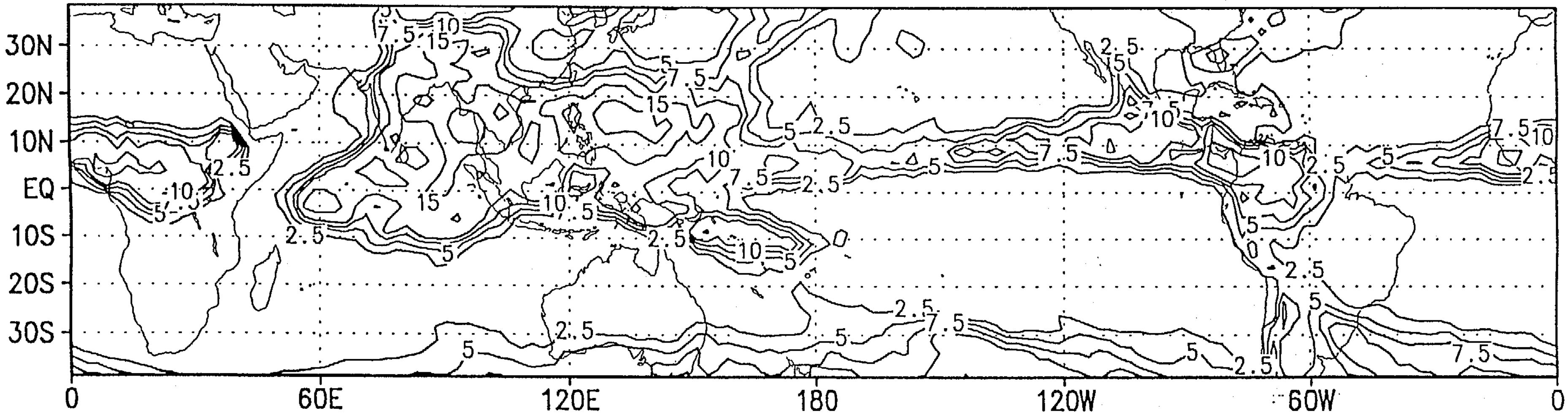
nosat

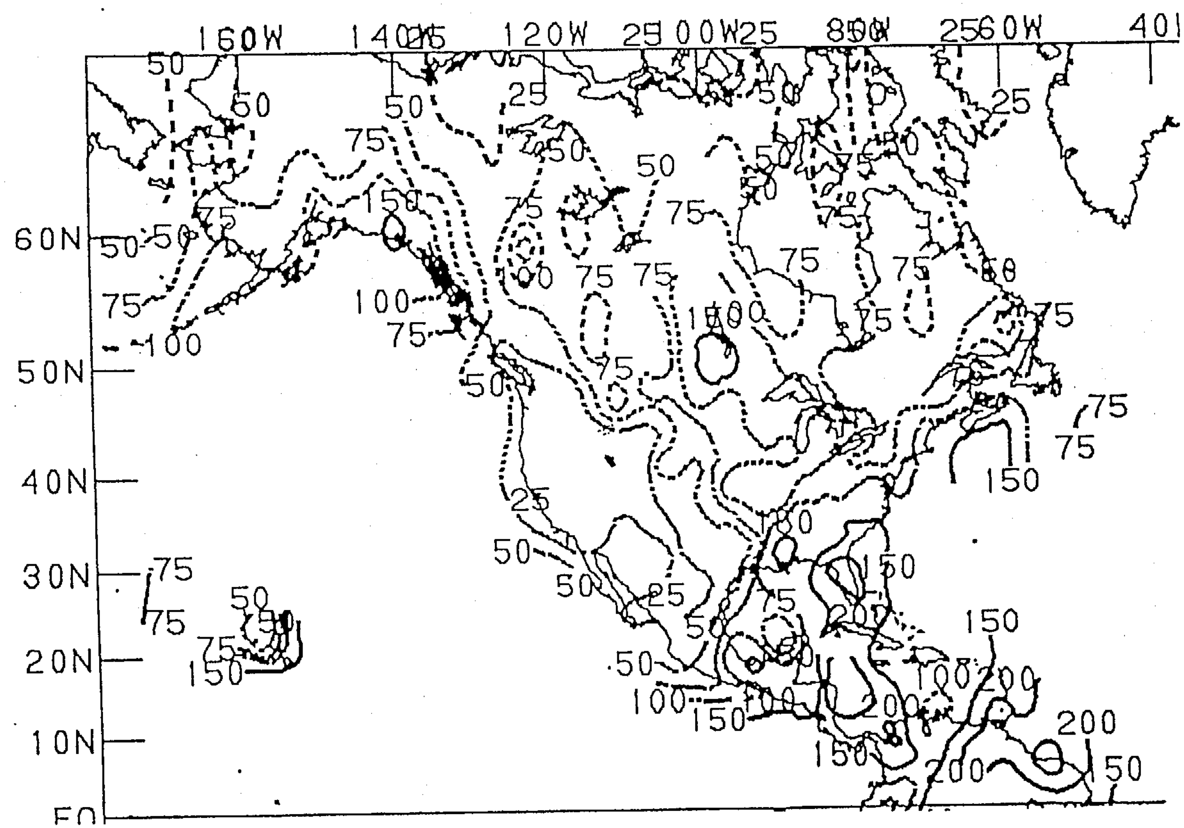


3d

GrADS: COLA/UMCP

GPI 8508





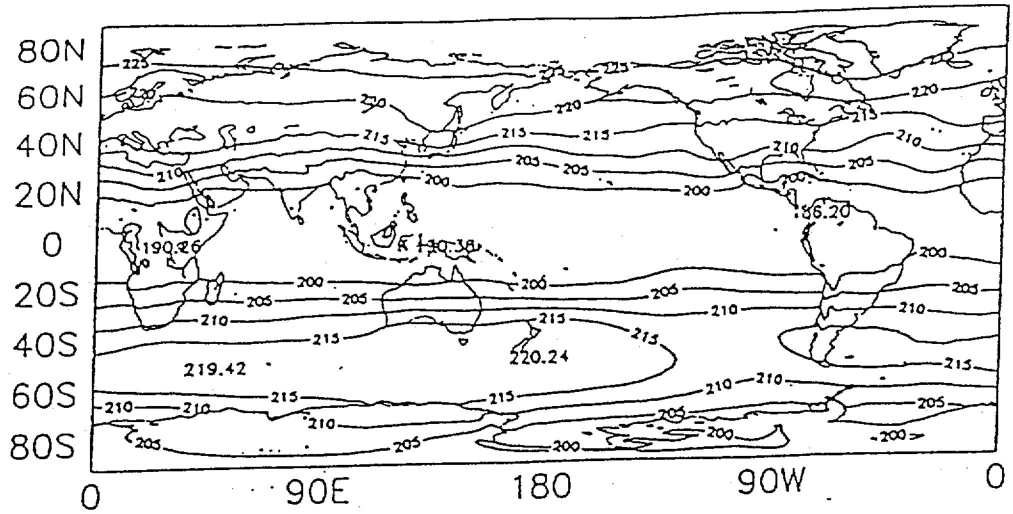
IX3

3f

6

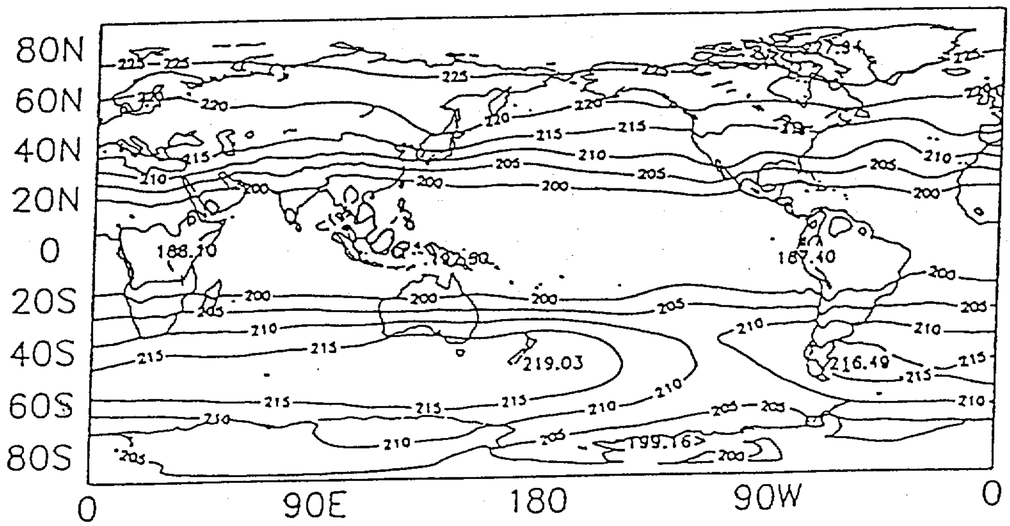
IX4

T126



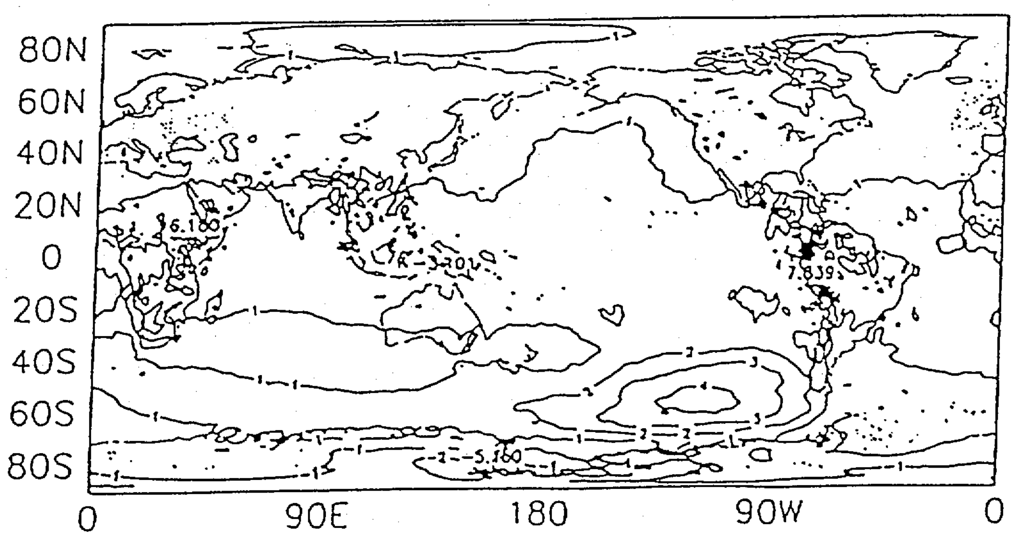
4a

T62



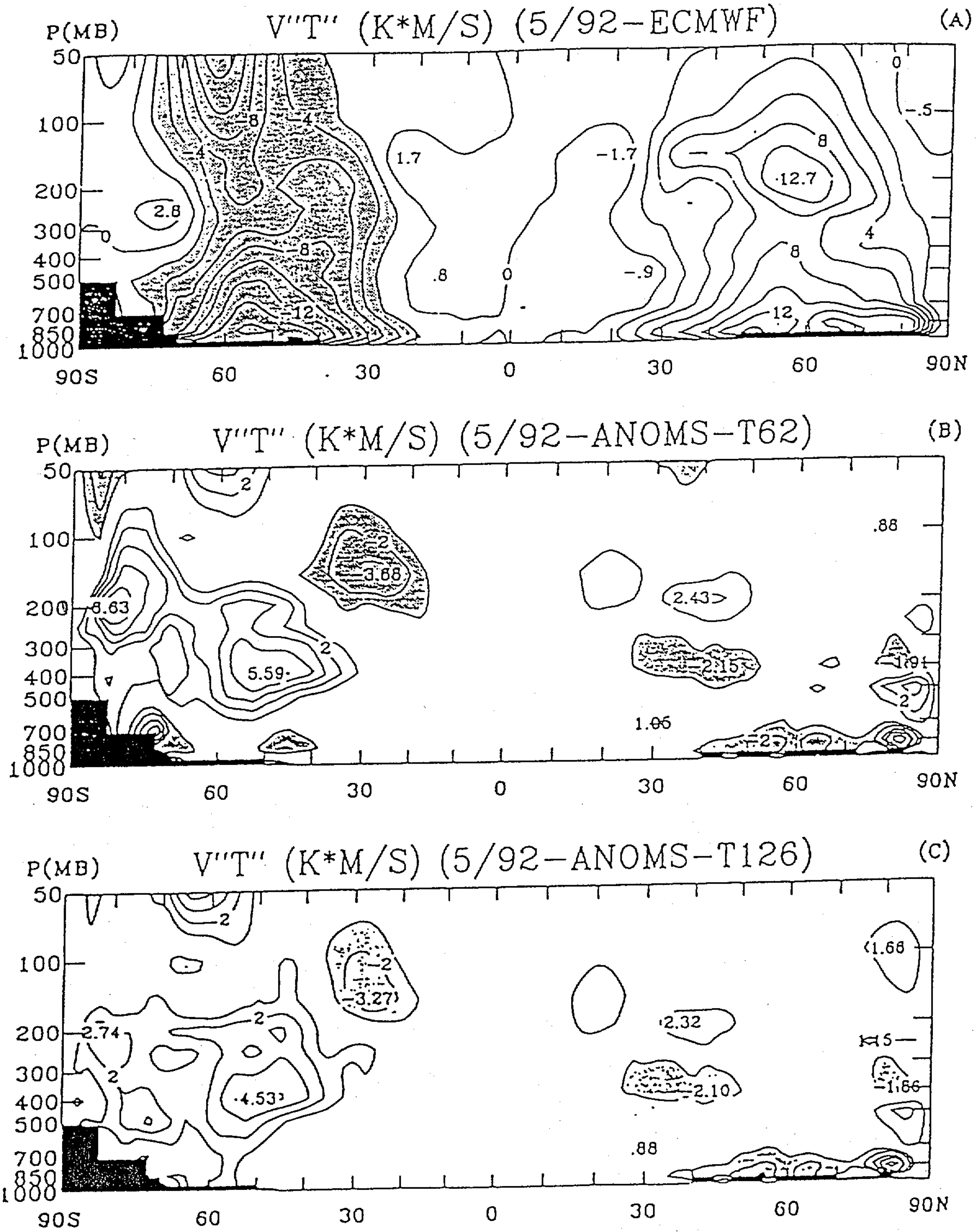
4b

T126 - T62



4c

IX.5



V''T'' (K*M/S) (5/92-ECMWF)
(A)
P(MB)
90S
60
30
0
30
60
90N
V''T'' (K*M/S) (5/92-ANOMS-T62)
(B)
V''T'' (K*M/S) (5/92-ANOMS-T126)
(C)

a

b

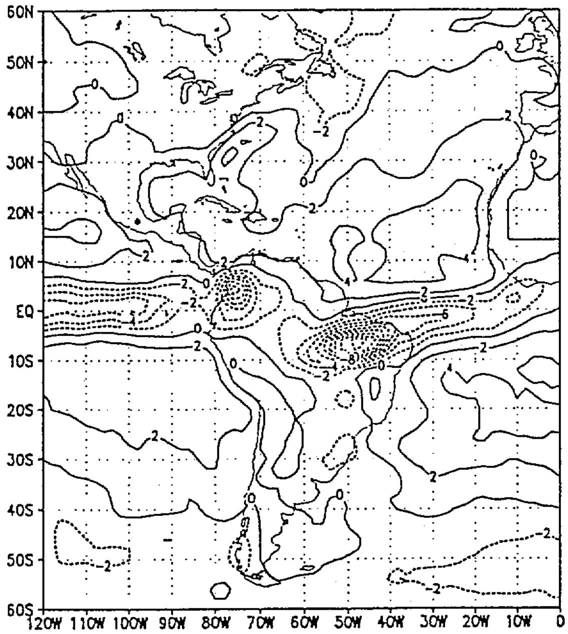
c

IX.6

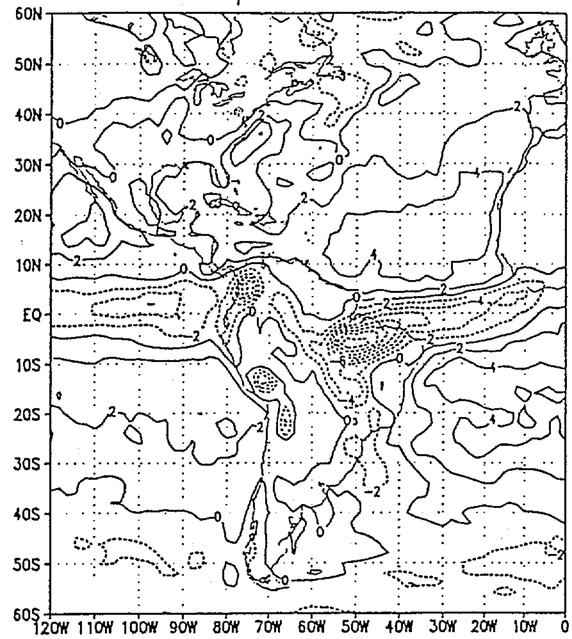
a

b

E-P T126



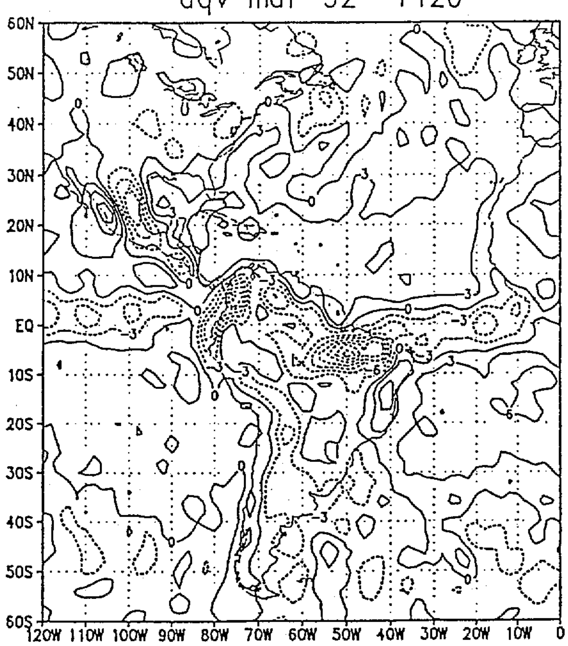
e-p t62 mar92



a

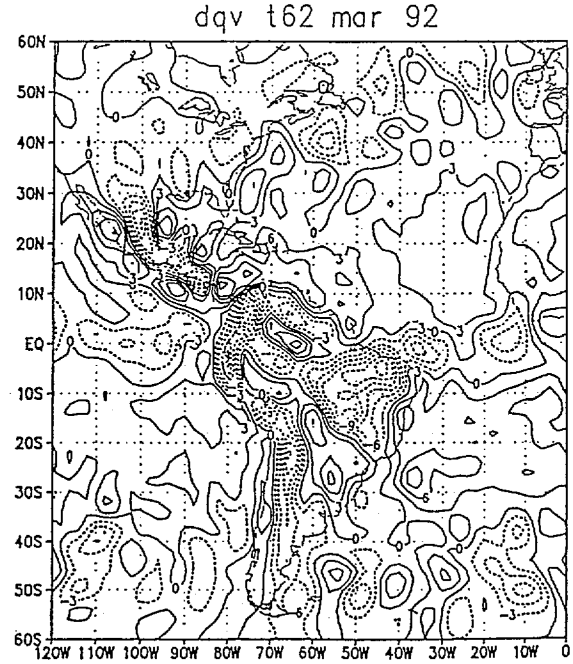
c

dqv mar 92 T126



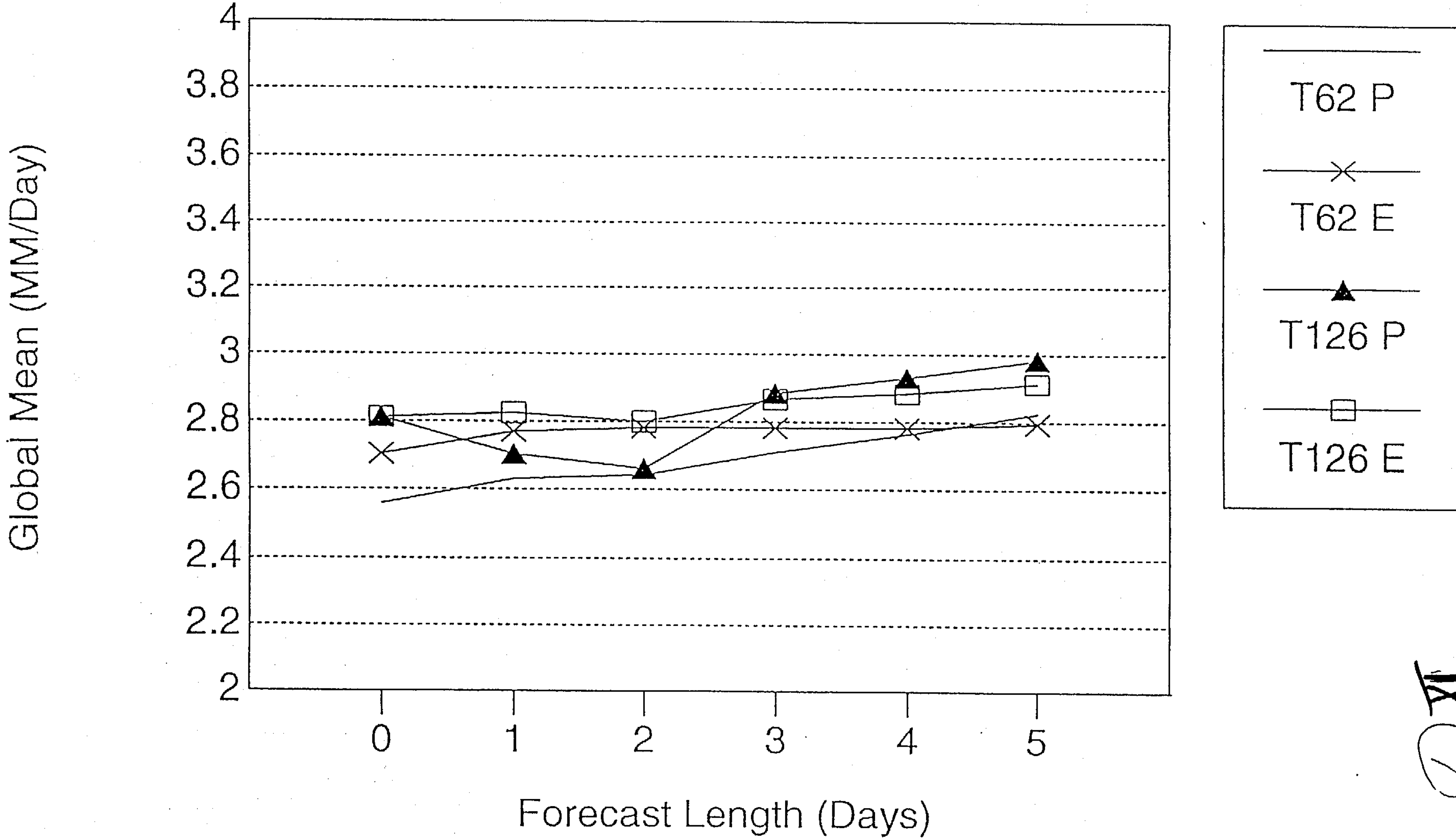
d

dqv t62 mar 92



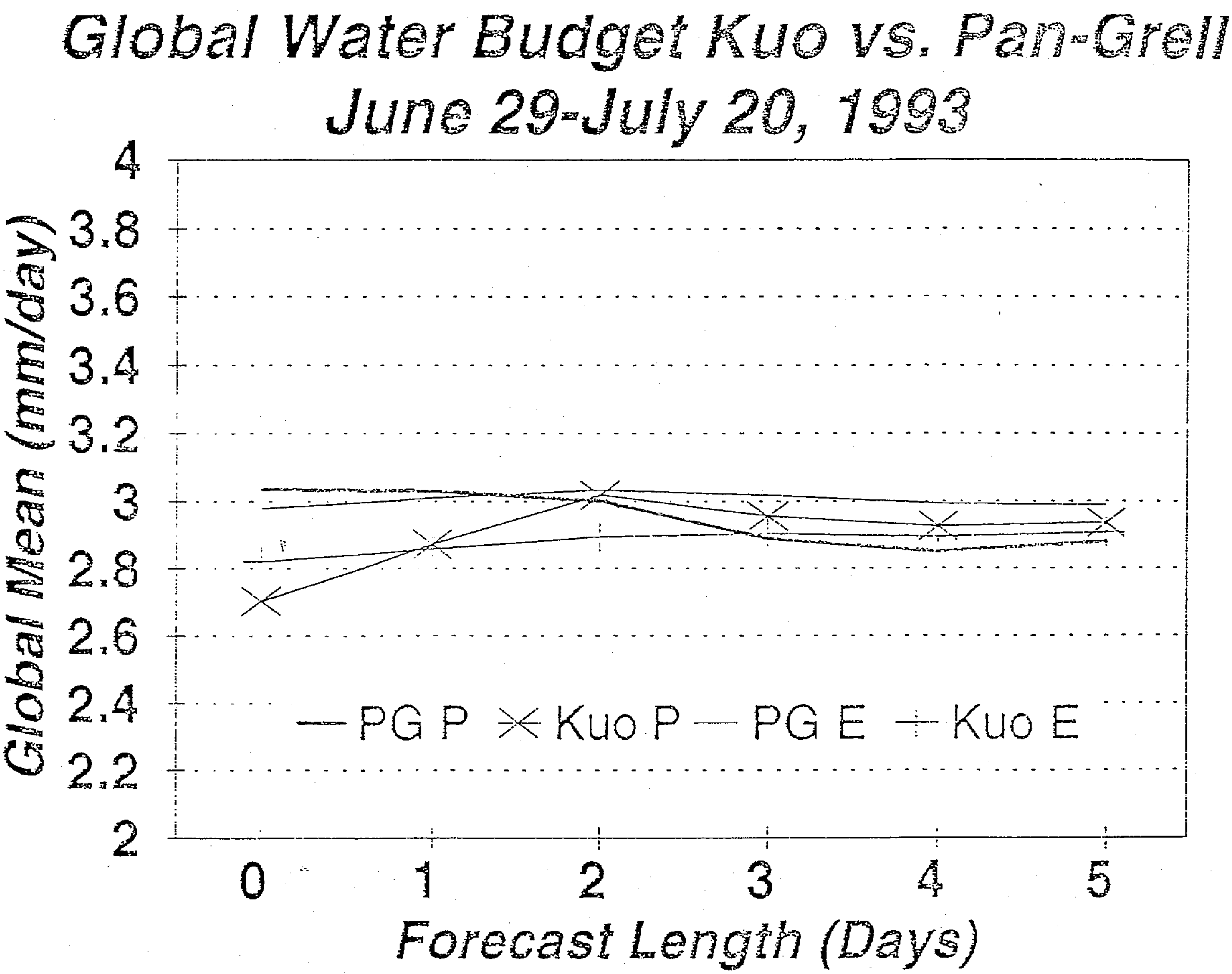
Global Water Budget

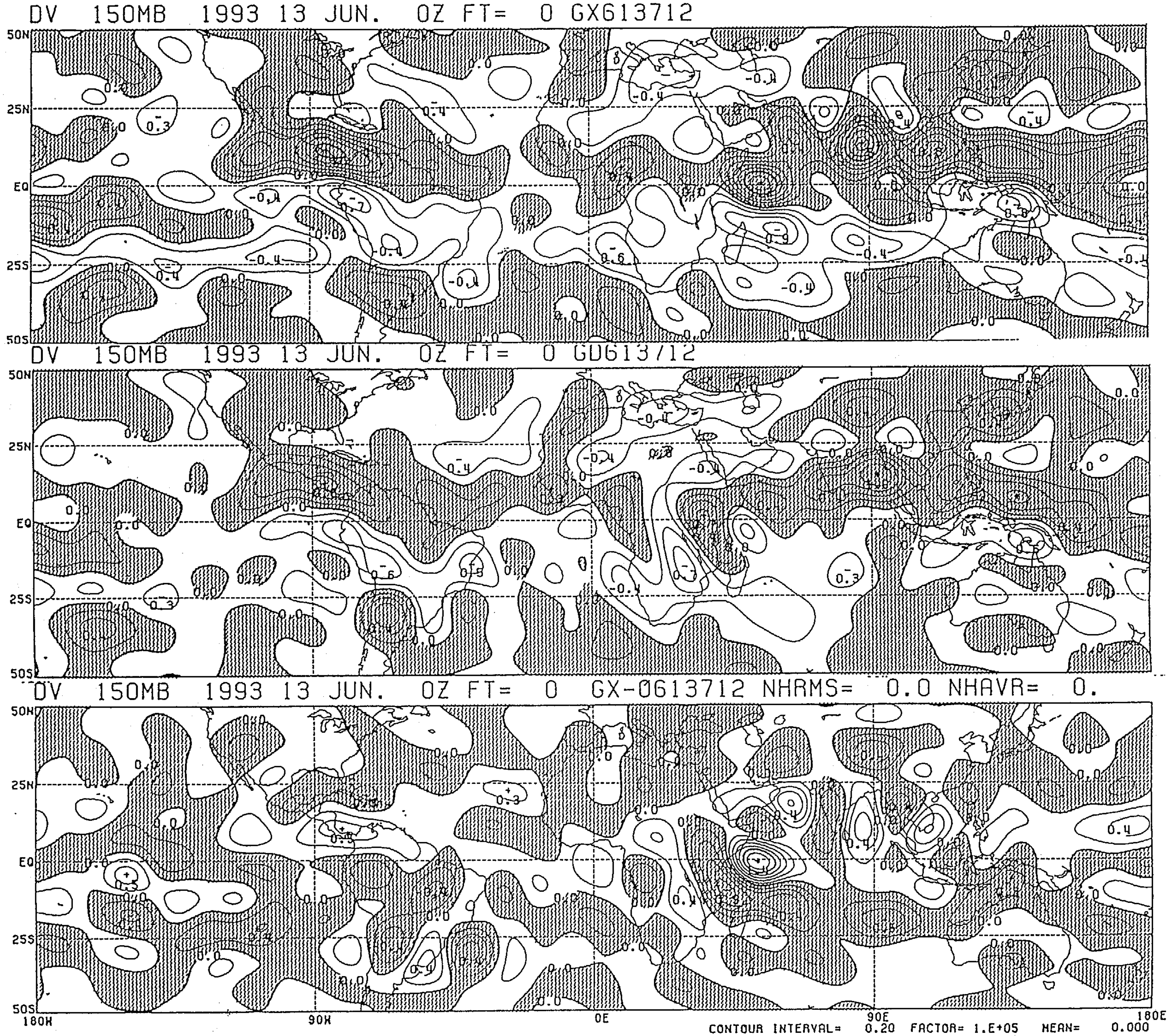
April-Sept. 1992

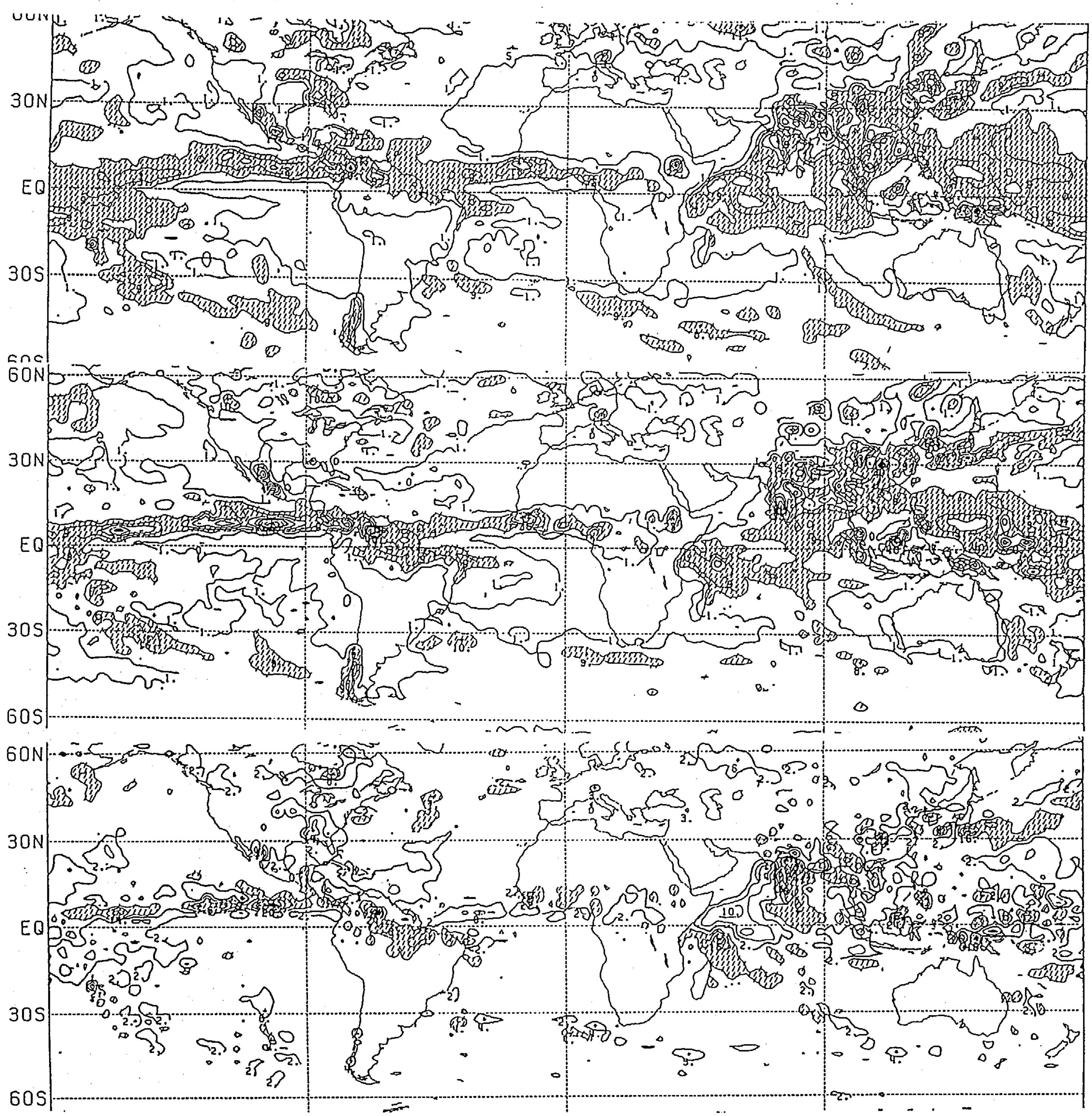


XII

IX 8

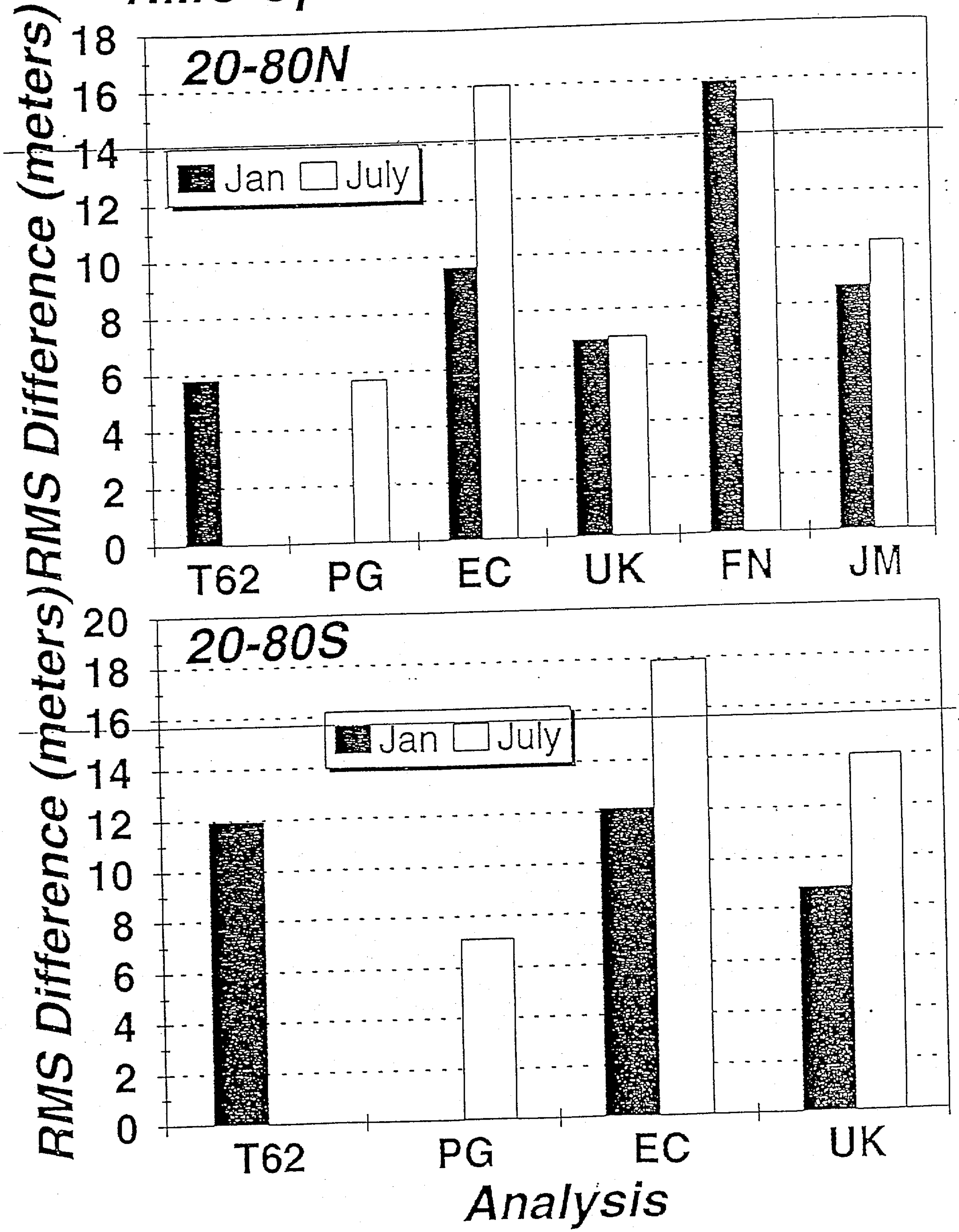

Global Water Budget Kuo vs. Pan-Grell
June 29-July 20, 1993
Global Mean (mm/day)
4
3.8
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
PG P
Kuo P
PG E
Kuo E
0
1
2
3
4
5
Forecast Length (Days)


DV 150MB 1993 13 JUN. 0Z FT= 0 GX613712
DV 150MB 1993 13 JUN. 0Z FT= 0 GD613712
DV 150MB 1993 13 JUN. 0Z FT= 0 GX-0613712 NHRMS= 0.0 NHAVR= 0.
50N
25N
EQ
25S
50S
180W
90W
0E
90E
180E
CONTOUR INTERVAL= 0.20 FACTOR= 1.E+05 MEAN= 0.000


60N
30N
EQ
30S
60S
60N
30N
EQ
30S
60S
60N
30N
EQ
30S
60S

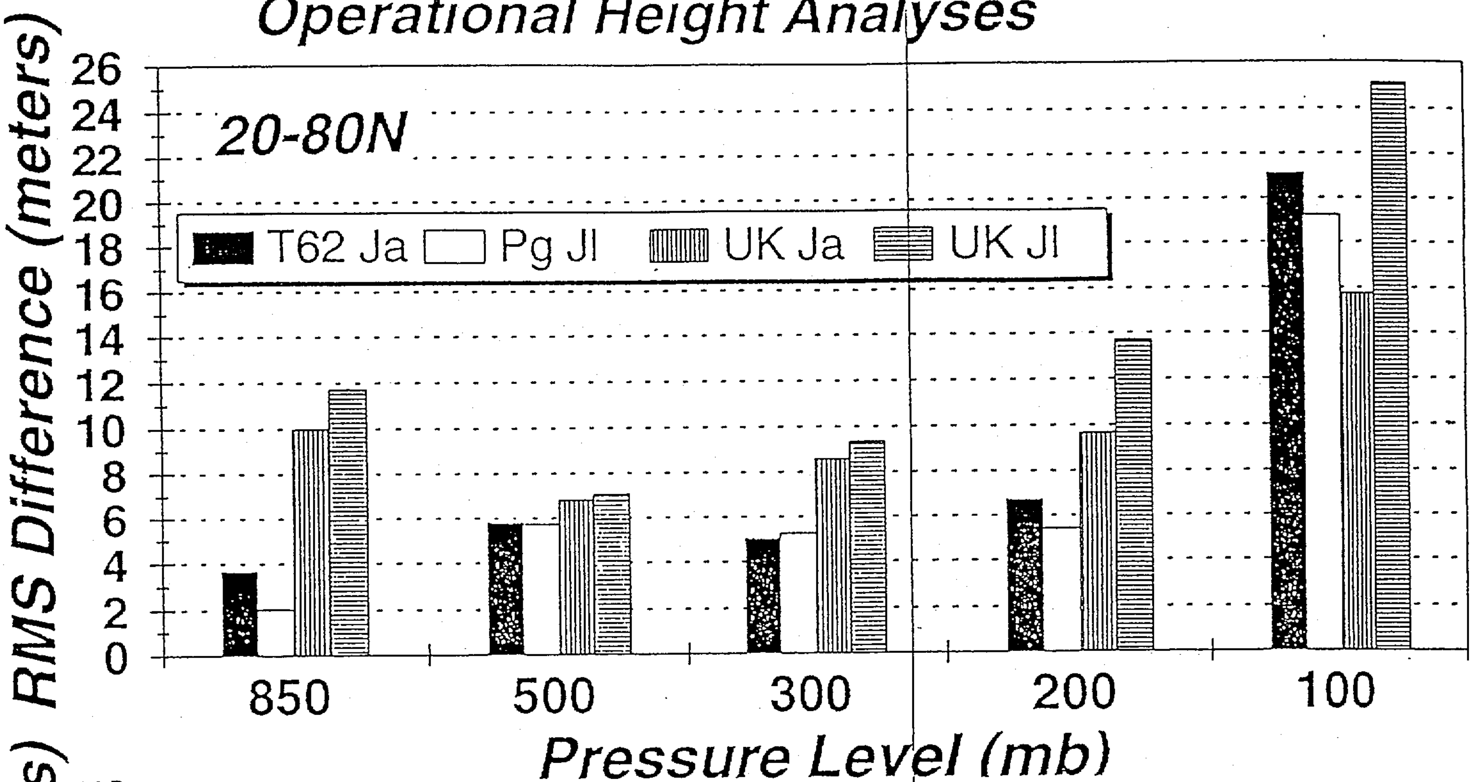
tx 11

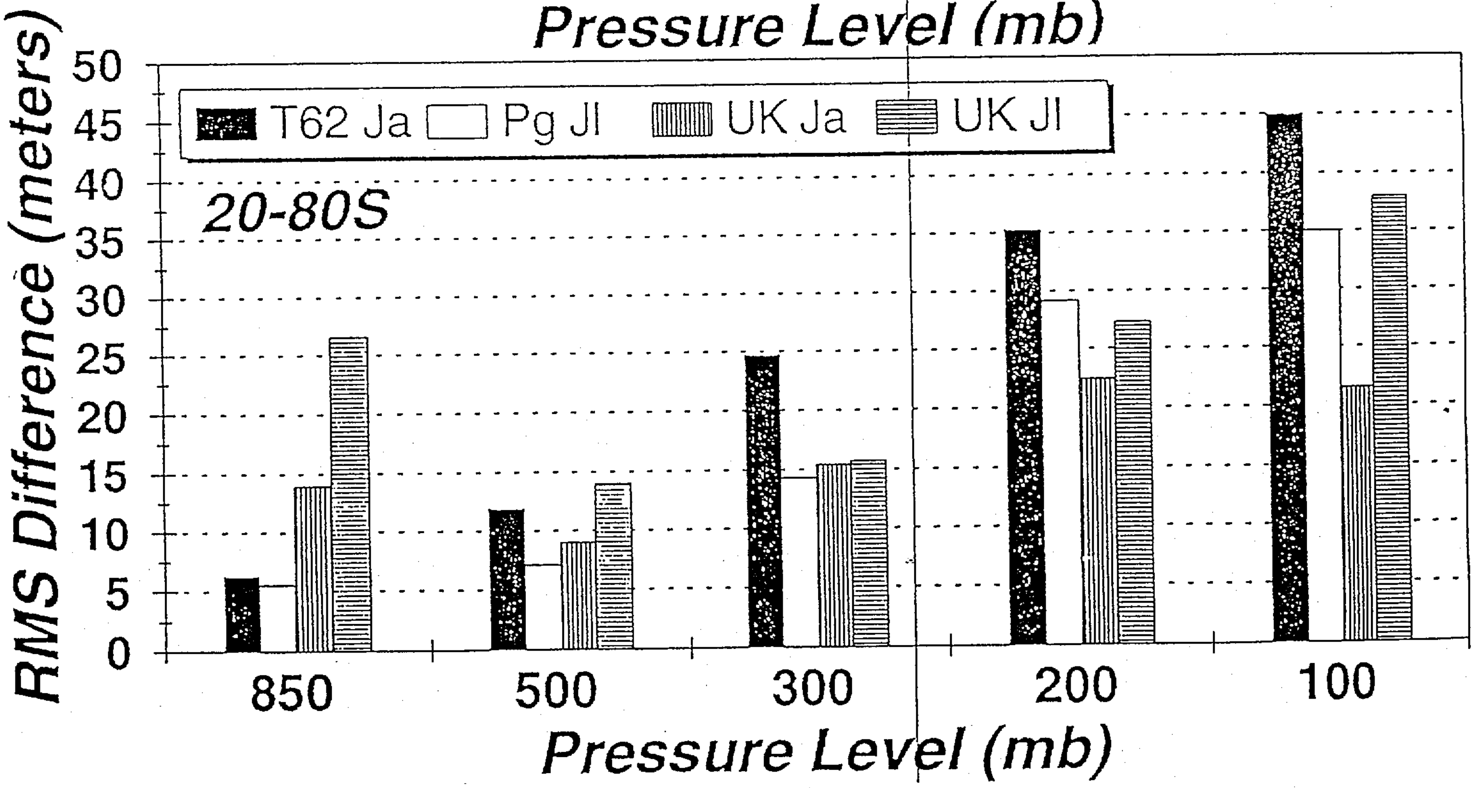
Difference from Monthly Mean Z500
NMC Operational GDAS

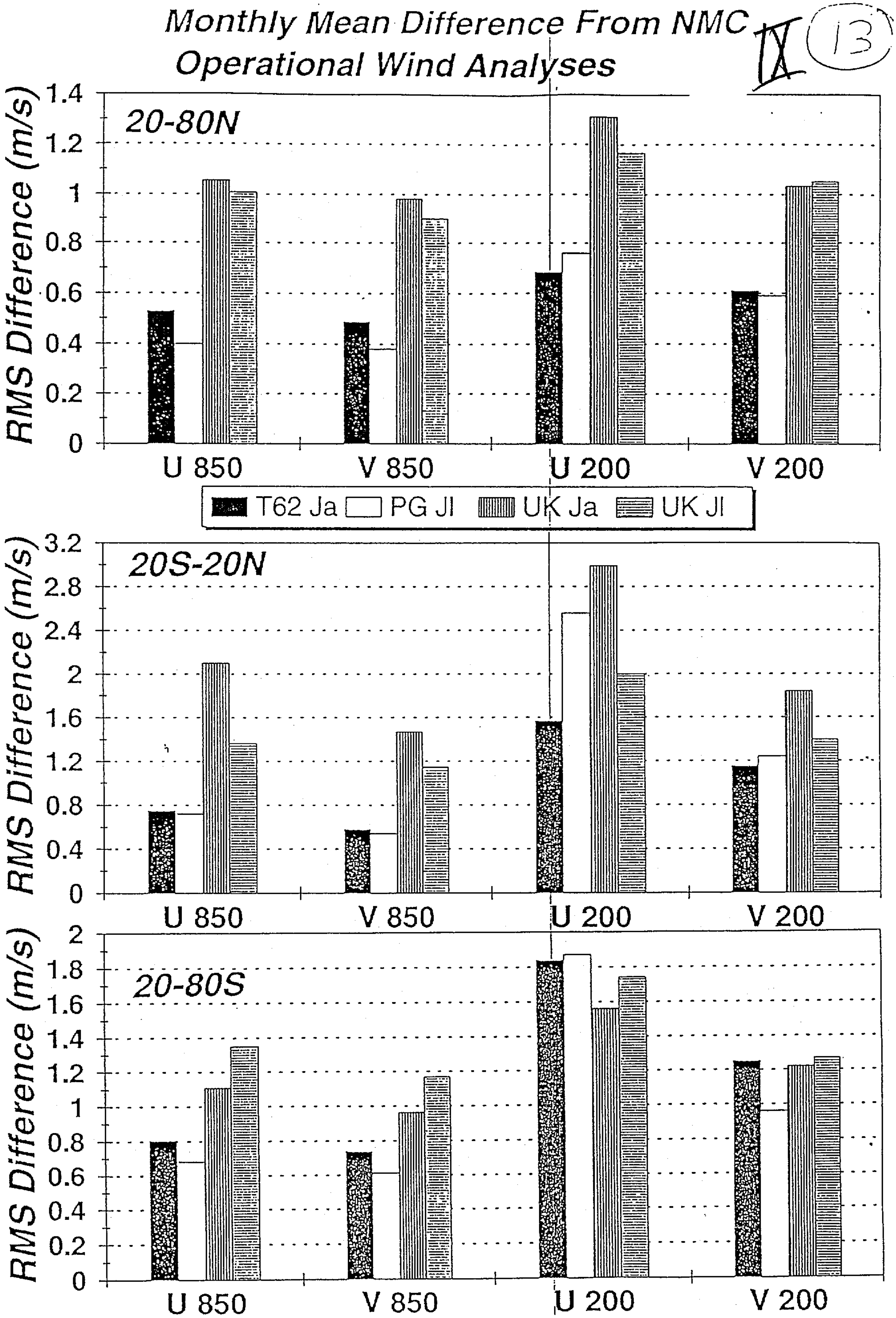


12

Monthly Mean Difference From NMC Operational Height Analyses







Monthly Mean Difference From NMC
Operational Wind Analyses
20-80N
20S-20N
20-80S
RMS Difference (m/s)
U 850
V 850
U 200
V 200
T62 Ja
PG Jl
UK Ja
UK Jl